Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12925 p-ISSN : 2541-044X

Improving IT Support Efficiency Using AI-Driven Ticket Random Forest Classification Technique

Nathaniel Crosley[1)*], Ito Wasito[2)]
[1,2,)]Universitas Pradita, Tangerang, Indonesia
[1)]nathaniel.crosley@student.pradita.ac.id, [2)]ito.wasito@pradita.ac.id

Submitted : Aug 22, 2023 | **Accepted** : Aug 26, 2023 | **Published** : Oct 1, 2023

Abstract: This research project aims to improve IT support efficiency at Indonesian company XYZ by using AI-based IT support ticket classification integration. This method involved collecting over 1,000 support tickets from the company's IT ticketing system, GLPI, and pre-processing the data to ensure the quality and relevance of the data for analysis. Claims data is enriched with relevant features, including textual information and categorical attributes such as urgency, impact, and requirement expertise. To improve the ticket description understanding, AI-based language models, especially IndoBERT, are used. These will help to reclassify and improve the work of IT support teams. In addition, the ticket data is used to train the Random Forest classifier, allowing automatic classification of tickets based on their specific characteristics. The performance of the ticket classification system is evaluated using a variety of metrics, and the results are compared with alternative methods to assess effectiveness of the Random Forest algorithm. This evaluation demonstrates the system's ability to correctly classify and prioritize incoming tickets. The successful implementation of this project at Company XYZ is a model for other organizations looking to optimize their IT support through AI-driven approaches. By providing simplified ticket classification and admission ticket reclassification based on AI algorithms, this research helps leverage AI technologies to improve IT support processes. Ultimately, the proposed solution benefits both support providers and users by improving efficiency, response times, and overall customer satisfaction.

Keywords: AI Integration, Efficiency, IT Support, Random Forest, Ticket Classification

INTRODUCTION

Effective troubleshooting and prompt resolution of user issues are paramount in the IT support sector(Ali, 2018). The rapidly increasing complexity of IT services in enterprises necessitates advanced systems for IT ticket classification(Revina et al., 2020). However, it has been found that IT support teams frequently encounter challenges in categorizing incidents correctly and routing them to the right resolution group(Silva et al., 2018). Such obstacles can result in longer search times, extended resolution intervals, and decreased user satisfaction (Brüggen et al, 2021). A key component of these challenges is the vast diversity of e-learning content available, making it challenging for learners to find suitable materials based on their knowledge levels(Thomas & J., 2020).

In light of these pressing challenges, the application of artificial intelligence (AI) techniques in various domains(Biau & Scornet, 2015), including paper currency classification and Big Data analytics,

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12925
e-ISSN : 2541-2019
p-ISSN : 2541-044X

has gained traction (Herrera et al., 2019). The foundational importance of a comprehensive knowledge base in supporting such AI implementations cannot be overstated. Technicians depend on it for swift access to pertinent troubleshooting information (Dai, 2022). Modern AI-powered models can produce detailed content for knowledge bases, covering various troubleshooting scenarios(Thomas & J., 2020), thereby simplifying the ticket categorization process (Anwar, 2021).

However, even with advanced systems, challenges persist. While the implementation of IT software or tools can help manage incoming tickets, there are still bottlenecks to be fixed (Qamili et al., 2018). The correct implementation of algorithms like Random Forest in high-performance computing platforms can also further enhance these processes (Herrera et al., 2019). Moreover, the potential of AI-enabled service chains, which requires the alignment of Service Level Agreements (SLAs) with AI systems, signifies the next frontier in optimizing operational efficiency (Engel et al., 2022). Research suggest with the advancement of AI privacy should also be focused (Elliott & Soifer, 2022), therefore the usage of data during experiment will be redacted based on enterprise security and privacy.

By leveraging these AI-based tools and models, technicians can access data that's not only accurate but also contextually relevant to the enterprise. The prioritization matrix are created to help with multiple bottlenecks in time management, easy task which takes time to complete (Kc et al., 2017) and the management of prioritization matrix itself (Bugayenko et al., 2023; Obodo, 2018). To tackle this problem simple 3x3 matrix is created based on Impact and Urgency of the ticket based on the ticket description. The result, is to create a models that understand the priority in which should be done now or later.

Furthermore, as the adoption of AI in various sectors such as business governance (Schneider et al., 2023) and public administration continues to grow, so does the need for continuous research on its potential implications and challenges.

LITERATURE REVIEW

The IT support field is constantly adapting to meet technological developments and the need to improve user interactions. Machine learning, with a particular focus on the Random Forest algorithm, plays an important role in these adjustments. In this literature review, we analyze relevant studies, detailing their approach, results, and broader impact. This systematic review aims to shed light on both progress and areas for improvement, paving the way for a clearer understanding of current research gaps.

Table. 1 Previous Studies which discuss IT Supports and Random Forest Classification

Author	Topic	Advantage	Disadvantage
(Ali, 2018)	Automatic Complaints Categorization Using Random Forest and Gradient Boosting	Effective categorization techniques for customer complaints using machine learning.	Specific focus on customer complaints, which might not encompass the entire breadth of IT support issues.
(Engel et al., 2022)	SLA-aware operational efficiency in AI-enabled service chains: challenges ahead	AI-driven knowledge base expansion integrated with Random Forest led to operational efficiency improvements in service chains.	The research might not fully account for unexpected challenges arising from integrating AI and SLA in dynamic service environments.
(Katuwal & Suganthan, 2018)	Enhancing Multi-Class Classification of	Combining Random Forest with other ML	Limited to multi-class problems; the

*name of corresponding author

	Random Forest using RVFNN and Oblique Decision Surfaces	techniques led to enhanced multi-class classification accuracy.	adaptability to real-world diverse and evolving issues remains uncertain.
(Paramesh et al., 2018)	Classifying Unstructured IT Service Desk Tickets Using Ensemble of Classifiers	High accuracy in classifying unstructured IT service desk tickets using ensemble classifiers, including Random Forest.	The research might be highly tailored to specific datasets, potentially limiting broader applicability.
(Dai, 2022)	Improving Random Forest Algorithm for University Academic Affairs Management System Platform Construction	Enhanced prediction accuracy and stability in academic affairs management systems.	Very niche application to university support services; might not be universally applicable to broader IT support domains.
(Gupta et al., 2018)	Reducing user input requests to improve IT support ticket resolution process	Significant reduction in user input requests; preemptive system demonstrated a high accuracy of 94-99% in predicting information needs, outperforming traditional methods.	The study focused on a specific aspect of the ticketing system, which might not encompass all facets of IT support. Further, it was case-specific.

The table above provides insight into the role of machine learning in resolving IT support requests, with particular emphasis on the capabilities of the Random Forest algorithm, among other techniques. While these methods are effective, they also pose operational challenges and areas for further research. Notable research gaps include the search for systematic solutions to prioritizing dynamic requirements, the need for models that adapt quickly to changing computing conditions, the potential benefits of combines various AI-based solutions and the mission-critical problem of ethical issues, such as manipulating SLA clocks, in the burgeoning IT-enabled field of AI. These areas will inform our future research directions to improve and enhance IT support activities.

METHOD

This section presents the framework of methodology used to improve the performance of XYZ Indonesia's IT support system. The central goal is achieved through automated requirement prioritization frameworks, using measurable impact and urgency metrics to establish a structured priority matrix.

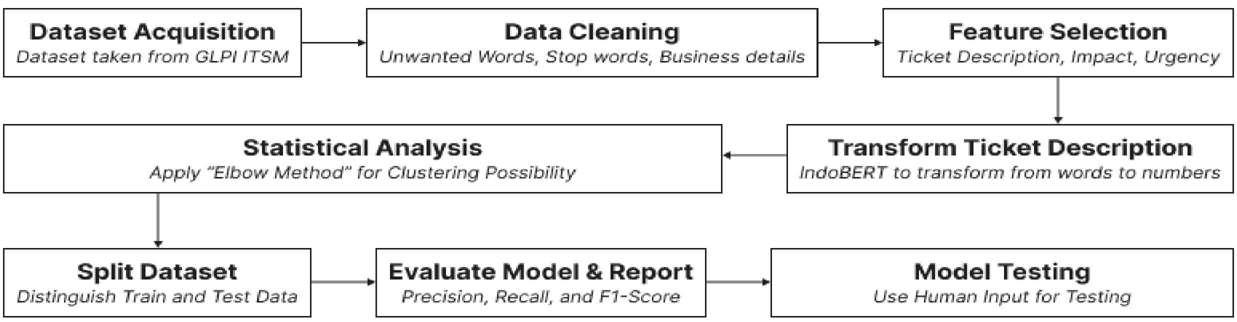

Fig. 1 Research Workflow
Source: researcher property

Methodological choices are guided by information gathered through collaboration with IT professionals within the organization. These experts have valuable expertise in this area and have been consulted to

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12925
e-ISSN : 2541-2019
p-ISSN : 2541-044X

provide information on the selection of appropriate methods. Their input helped determine which analysis tool is best suited for developing an automated request prioritization framework. The basis of the analysis is data from the IT ticketing system GLPI, spanning from October 2022 to June 2023. This dataset includes a set of supporting tickets as the basis for the extent investigation.

Extensive data pre-processing is performed to resolve any inconsistencies or gaps in the data. This step is intended to remove critical business related information and to ensure the reliability of subsequent analyses. Feature extraction is performed post pre-processing the dataset, this includes three selections, Urgency, Impact, and Ticket descriptions. To extract relevant information from these tickets, the IndoBERT model was used. This model, a variant of the BERT architecture adapted to the Indonesian language, is the engine that encodes the textual data of the ticket, converting it into a format suitable for model import. Once encoded, the IndoBERT model generates digital representations that encapsulate the context of textual data. These integrations become necessary for the subsequent classification process.

The main element of this study is the 3x3 matrix, which is built on the basis of two main dimensions: Impact and urgency. The Impact dimension has been subdivided into:

1. High: Pertaining to issues with organizational-wide repercussions.
2. Medium: Referring to challenges localized to particular departments.
3. Low: Denoting concerns with individual implications.

On the other hand, the Urgency dimension was demarcated as:

1. High: Demanding instantaneous resolution.
2. Medium: Necessitating resolution within a day.
3. Low: Issues that can be addressed with a broader resolution window.

PRIORITY IMPACT * URGENCY		**IMPACT**		
		Low Personal Level	**Medium Department Level**	**High Organizational Level**
URGENCY	**Low Later date**	1	2	3
	Medium Within 24 hours	2	3	6
	High Immediately / Now	3	6	9

Fig. 2 Prioritization Matrix
Source: researcher property

By cross-referencing aspects of Impact and Urgency, the matrix provides a systematic approach to prioritizing issues. The priority will be scaled from 1 to 9 in which empower the Urgency levels:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12925
e-ISSN : 2541-2019
p-ISSN : 2541-044X

1. Low (1 - 2): Can be done at later date.
2. Medium (3 - 6): Should be done within 24 hours, this might impact organization at scale.
3. High (9): This should be urgent and has impact on the whole organization/enterprise.

With integrations native to IndoBERT, this experience is trained by the Random Forest classifier to recognize and understand patterns that associate request content with their respective Impact and Urgency classifications. . After the model is properly trained, it is used to predict the ticket classification in the test subset, based on the established matrix.

The next stage is devoted to evaluation. This study uses a subset of tests to evaluate classifier performance, applying a series of measures such as confusion matrix, precision, accuracy, recall, and F1 scores. This assessment makes it possible to quantify the model's predictive power, by highlighting areas of accuracy and potential discrepancies.

At the end of the evaluation, the tickets are systematically classified according to a 3x3 matrix. The purpose behind this automatic classification is to provide a methodical approach for IT support staff to handle requests, ensuring that the most critical issues are resolved quickly.

RESULT

Dataset is taken from the GLPI system through the use of an export to csv function. The dataset used during the research process is modified to use all the available columns which are available in the GLPI system. This to ensure that the data is available to use.

In the data preprocessing stage, data cleaning was performed to ensure the security and relevance of the data set. Specifically, this process is intended to eliminate company-specific details that could serve as potential identifiers, such as personal names, email addresses, affiliations with specific organizations or departments. and exact location of buildings. . This process is integral to maintaining the security of the ticket issuer while preserving the nature of the data for analysis.

From the large collection of data obtained, the analysis focused on a subset of 1,000 ticket descriptions. This decision is strategic: The subset is organized to include commonly encountered requirements and handled by IT professionals across the enterprise, ensuring that the model is trained on representative and relevant data. Features play a central role in machine learning applications, determining model accuracy and reliability. For this study, the main characteristics extracted from the cleaned data set were the Urgency and the Impact associated with each ticket as well as the "text description of the ticket". The first two serve as parameters denoting the priority of the ticket, while the second provides qualitative information about the nature of the request or IT problem.

To convert the text descriptions into a machine-readable format, the first tokenization stage was used, which split the text into individual tokens or words. These tokens are then encrypted using the IndoBERT model. Text descriptions are encoded using three tensors, input_ids which represent words or subwords, attention_mask, and token_type_ids which differentiates between multiple tasks. This data will then be fed to Random Forest to absorb.

Regarding the evaluation of existing priorities, it was found that the current ticket classification is mainly based on human judgment. In contrast, this study aims to provide a more algorithmic approach: The priority is calculated from the product of the Impact and Urgency values, as predicted by the trained model.

At the beginning of the study, the aim was to implement a 5x5 preference matrix which provides more granular information for Impact and Urgency. However, a detailed analysis of the data set shows that

*name of corresponding author

its structure is limited to 3x3 matrices. To validate this limitation, cluster analysis was applied, in particular using the "Elbow Method". This technique aims to determine the optimal number of clusters by testing the variance against the number of clusters. The results clearly indicate the presence of three clusters, supporting a 3x3 matrix configuration. Based on this finding, adjustments to the study method were made.

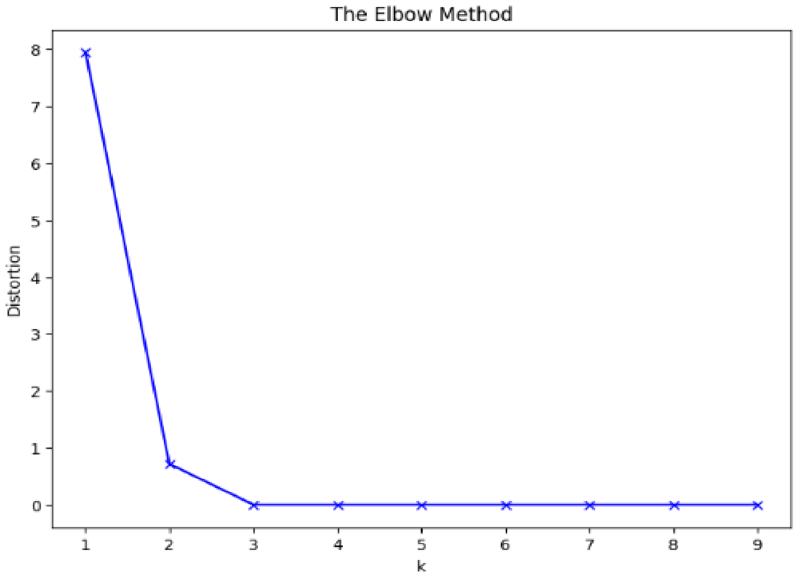

Fig. 3 Elbow Method
Source: researcher property

With the revised approach, the focus shifts to reclassifying data points, especially those that are not defined as "high", "medium" or "low" in terms of impact and emergencies. For this task, the Random Forest classifier was used due to its ability to handle multidimensional classification challenges. The reclassification is informed by the patterns seen in the 3x3 matrix.

Following the data split, the training dataset serves as the basic data source for the Random Forest models to learn and adjust. The test dataset, on the other hand, remains untouched during training, solely to evaluate the model's accuracy and performance on unseen data. This descriptor is necessary to maintain the integrity of the evaluation process and to assess how generalizable the models are to new, unseen data.

The decision to use 100 decision trees in the Random Forest model is made based on a balance between computational efficiency and prediction accuracy. Too few trees may not capture all the subtleties and patterns of the data, while too many trees may lead to reduced returns in terms of performance improvements and increased computational costs. To ensure consistent results between different analyzes and possibly between different investigators, a fixed randomization state was established. This setting ensures that the randomness inherent in the Random Forest algorithm, such as sampling and feature selection, remains consistent. This is an important step in scientific research as it ensures that experiments and results can be reproducible.

Then, separate random forest models, one for impact and another for urgency, are then trained on the training dataset. These models work by taking advantage of the collective decision-making power of each decision tree. Each tree makes predictions based on the provided features, and the final model outcome is determined by aggregating these individual predictions. In the context of this study, the objective was to determine the extent to which these models were able to predict the impact and urgency of the IT support ticket based on features extracted from the ticket data. request.

*name of corresponding author

During the evaluation phase, key metrics are used to evaluate the performance of the model. The confusion matrix presents a breakdown of the classification results, while the accuracy index quantifies the percentage of correct predictions. Given the unequal distribution of the dataset, measures such as precision, recall, and F1 score were included to provide a more comprehensive assessment. Although other measures, such as ROC and AUC curves, may have been taken into account, the focus remains on the primary measures. In addition, further examination of the importance of features in the Random Forest classifier may be pursued in future research projects.

DISCUSSIONS

Based on the earlier result, the prioritization matrix that will be used is based on 3*3 matrix. Below are the breakdown of each metrics.

1. **High**, marked as '3': This level is for the most pressing tickets that need immediate action.
2. **Medium**, marked as '2': Tickets at this level are important but not super urgent. They should be addressed after the high-priority ones.
3. **Low**, marked as '1': These tickets aren't urgent and can wait a bit longer to be addressed.

This matrix will be used for detailing further analysis of the research.

Urgency Metrics Analysis

The results for the urgency metrics shed light on the model's performance concerning the classification of ticket urgencies.

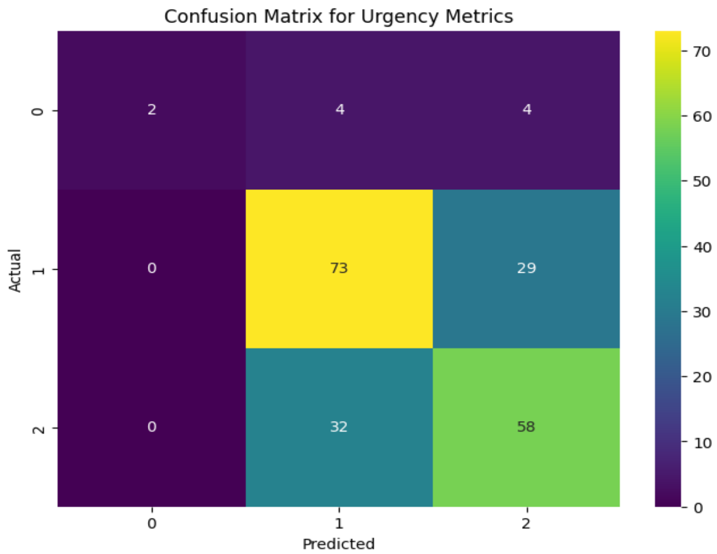

Fig 4. Confusion Matrix for Urgency Metrics
Source: researcher property

Confusion Matrix:
The confusion matrix for urgency (refer to Fig 4), provides an overview of the model's classification across different urgency categories:

1. Category '1' saw 2 true positives. However, 8 tickets, in reality of this category, were misclassified, primarily distributed between categories '2' and '3'.
2. For category '2', 73 tickets were correctly classified, but there was a noticeable misclassification with 29 tickets incorrectly predicted as category '3' and none as category '1'.
3. Category '3' had a balanced spread with 58 correct predictions, but 32 tickets were wrongly classified as category '2'.

Classification Report:
Analyzing the classification report (shown in Fig 5):

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i4.12925 p-ISSN : 2541-044X

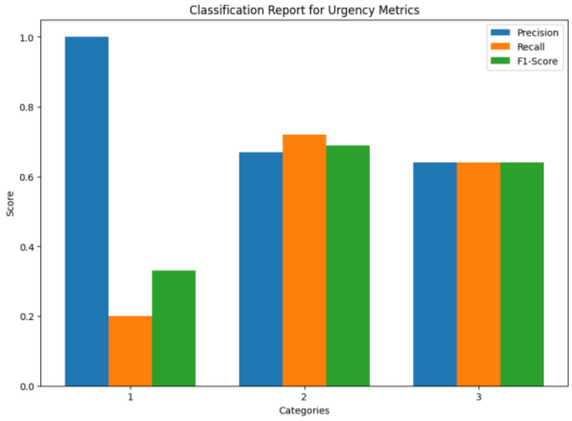

Fig 5. Classification Report for Urgency Matrix
Source: researcher property

1. Category '1' achieved perfect precision, meaning when the model predicted this category, it was always correct. However, the recall was significantly low at 0.20, indicating that out of all actual category '1' tickets, only 20% were correctly identified.
2. Category '2' had balanced precision and recall rates, both hovering around 0.7.
3. Category '3' demonstrated a symmetry between precision and recall, both approximating 0.64.

The overall accuracy of the model in predicting urgency stands at 0.66, signifying that two-thirds of the tickets were classified into their correct urgency categories.
2. Impact Metrics Analysis
The evaluation for the impact metrics provided insights into the model's proficiency in gauging ticket impact.

Confusion Matrix:
In the confusion matrix for impact (Fig 6):

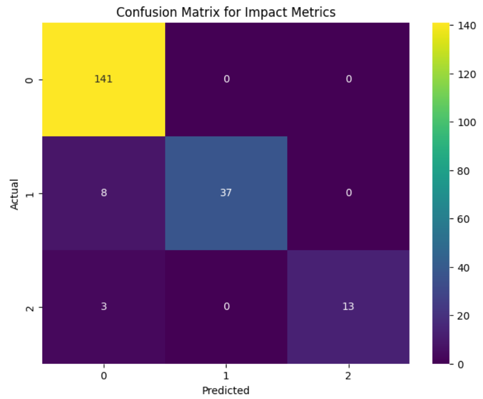

Fig 6. Confusion Matrix for Impact Metrics
Source: researcher property

1. Category '1' had an impeccable record with all 141 tickets correctly classified.
2. Category '2' had 37 tickets correctly identified. However, 8 tickets were incorrectly classified as category '1'. No tickets were misclassified as category '3'.
3. Category '3', the least represented category, saw 13 correct classifications but faced misclassifications with 3 tickets wrongly predicted as category '1'.

Classification Report:
Diving deeper into the classification report (Fig 7):

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12925
e-ISSN : 2541-2019
p-ISSN : 2541-044X

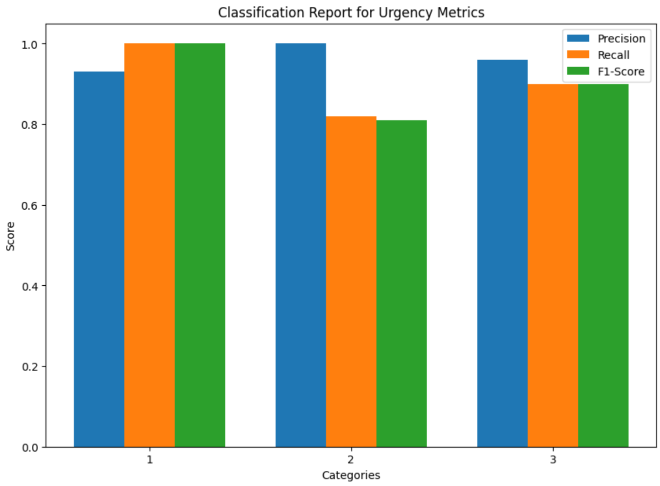

Fig 7. Classification Report for Impact Matrix
Source: researcher property

1. Category '1' showcased commendable precision and recall rates, with a striking perfect recall.
2. Category '2' stood out with a 1.00 precision, but the recall dropped to 0.82.
3. Category '3', despite its smaller sample size, boasted of high precision and recall, showcasing the model's efficacy in handling less frequent categories.

With an impressive overall accuracy of 0.95, the model demonstrated high reliability in predicting ticket impact.

Sample Predictions Analysis

Examining individual sample predictions provides a real-world perspective on the model's decisions:

1. For the description "Whatsapp tidak bisa diakses", the predicted urgency was '3' and the impact '1' "Printer tidak bisa digunakan" resulted in predictions of urgency '3' and impact '1'.
2. An "upgrade Pc" had urgency '1' and impact '1'.
3. Lastly, "Update MS.Office" was classified under urgency '2' and impact '1'.

These samples reiterate the model's tendency to lean towards certain urgency and impact categories, potentially due to the dataset's characteristics or the model's training nuances.

Visualization of Ticket Descriptions

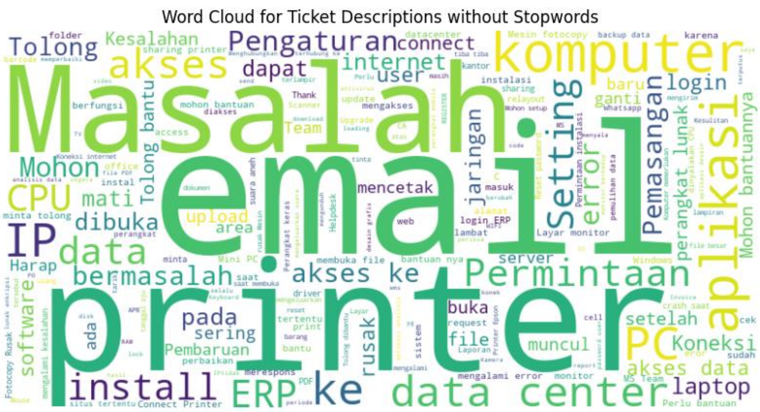

Fig 8. Word Cloud for Ticket Descriptions
Source: researcher property

A Word Cloud visualization (Fig 8) was generated to emphasize the most frequently occurring words in the ticket descriptions. This visualization helps in intuitively grasping the main concerns or features of

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 4, October 2023
DOI : https://doi.org/10.33395/sinkron.v8i4.12925
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the support tickets. Post the removal of Bahasa Indonesia stopwords, the word cloud accentuated keywords that play pivotal roles in the model's decision-making process.

There are several key takeaways from this study, from the consistency of the priority prediction that makes up the majority of typical support requests, to the fine-tuned data that will increase the accuracy of the models being evaluated. created after the test is complete. Although the matrix system provides a solid foundation for request prioritization, additional iterations and continuous training of the model can improve its accuracy. Future efforts may focus on expanding the dataset with a broader range of claim descriptions. Additionally, incorporating feedback from IT professionals can yield valuable insights to further refine the model. The ultimate goal is to move to a 5x5 matrix, which provides a more granular perspective that helps IT professionals understand the problem more effectively.

CONCLUSION

To enhance the prioritization of IT tickets, a 3x3 matrix based on urgency and impact has been implemented, offering a streamlined and effective approach to decision-making. This research employed advanced tools, including the IndoBERT model tailored for the Indonesian language and the Random Forest classifier, to assess the matrix's efficiency. While the results indicate considerable accuracy in specific domains, there remains room for further refinement in others.

REFERENCES

Ali, A. R. (2018). Cognitive Computing to Optimize IT Services.

Anwar, M. T. (2021). Automatic Complaints Categorization Using Random Forest and Gradient Boosting.
Advance Sustainable Science, Engineering and Technology, 3(1), 0210106. https://doi.org/10.26877/asset.v3i1.8460

Biau, G., & Scornet, E. (2015). A Random Forest Guided Tour (arXiv:1511.05741). arXiv. http://arxiv.org/abs/1511.05741

Bugayenko, Y., Bakare, A., Cheverda, A., Farina, M., Kruglov, A., Plaksin, Y., Pedrycz, W., & Succi, G. (2023). Prioritizing tasks in software development: A systematic literature review. PLOS ONE, 18(4), e0283838. https://doi.org/10.1371/journal.pone.0283838

Dai, J. (2022). Improving Random Forest Algorithm for University Academic Affairs Management System Platform Construction. Advances in Multimedia, 2022, 1–9. https://doi.org/10.1155/2022/8064844

Elliott, D., & Soifer, E. (2022). AI Technologies, Privacy, and Security. Frontiers in Artificial Intelligence, 5, 826737. https://doi.org/10.3389/frai.2022.826737

Engel, R., Fernandez, P., Ruiz-Cortes, A., Megahed, A., & Ojeda-Perez, J. (2022). SLA-aware operational efficiency in AI-enabled service chains: challenges ahead. Information Systems and E-Business Management, 20(1), 199–221. https://doi.org/10.1007/s10257-022-00551-w

Herrera, V. M., Khoshgoftaar, T. M., Villanustre, F., & Furht, B. (2019). Random forest implementation and optimization for Big Data analytics on LexisNexis's high performance computing cluster platform. Journal of Big Data, 6(1), 68. https://doi.org/10.1186/s40537-019-0232-1

Kc, D. S., Staats, B. R., Kouchaki, M., & Gino, F. (2017). Task Selection and Workload: A Focus on Completing Easy Tasks Hurts Long-Term Performance. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2992588

Obodo, N. A. (2018). Content Analysis of Time Management as a Tool for Corporate Effectiveness. International Journal of Applied Economics, Finance and Accounting, 2(2), 36–39. https://doi.org/10.33094/8.2017.2018.22.36.39

Qamili, R., Shabani, S., & Schneider, J. (2018). An Intelligent Framework for Issue Ticketing System Based on Machine Learning. 2018 IEEE 22nd International Enterprise Distributed Object Computing Workshop (EDOCW), 79–86. https://doi.org/10.1109/EDOCW.2018.00022

Revina, A., Buza, K., & Meister, V. G. (2020). IT Ticket Classification: The Simpler, the Better. IEEE Access, 8, 193380–193395. https://doi.org/10.1109/ACCESS.2020.3032840

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2023). Artificial Intelligence Governance For Businesses. Information Systems Management, 40(3), 229–249. https://doi.org/10.1080/10580530.2022.2085825

Silva, S., Pereira, R., & Ribeiro, R. (2018). Machine learning in incident categorization automation. 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), 1–6. https://doi.org/10.23919/CISTI.2018.8399244

Thomas, B., & J., C. (2020). Random forest application on cognitive level classification of E-learning content. International Journal of Electrical and Computer Engineering (IJECE), 10(4), 4372. https://doi.org/10.11591/ijece.v10i4.pp4372-4380

*name of corresponding author