

MACHINE LEARNING APPROACHES FOR HELPDESK TICKETING SYSTEM: A SYSTEMATIC LITERATURE REVIEW

ALVIAN SHANARDI WIJAYA¹, TANTY OKTAVIA²

^{1,2}Information System Management Department, BINUS Graduate Program, Master of Information System Management, Bina Nusantara University, West Jakarta, Jakarta, 11480, Indonesia

¹ alvian.wijaya@binus.ac.id, ² toktavia@binus.edu

ABSTRACT

Machine learning has been commonly used as a tool to support helpdesk function in many areas especially in ticketing system area. Ticketing system is one of the most common features for Helpdesk to provide support for users from answering question, resolving error and giving feedback to the products or services provided by the company. Combined with machine learning, helpdesk could classify user's problem according to their impact and urgency level by learning datasets produced helpdesk. This study aims to review relevant works about machine learning approaches in various use case scenario in helpdesk ticketing function. This study will perform a systematic literature review using PSALSAR Framework as a tool to study this knowledge based on SCOPUS database starting from 2012 until 2022. This paper will specifically search for three keywords ("Machine Learning", "Helpdesk" and "Ticketing") to find the related article and present a systematic literature review using three research question. Those findings will be discussed using quantitative, descriptive and narrative analysis to answer all of research question by manually assessed and extracting necessary data from each individual study This paper will produce systematical review of machine learning use cases and method in helpdesk ticketing function which could be useful for further research in other helpdesk ticketing function that has not been researched and provide analytical data about machine learning limitations and what could be done in the future to create more advanced machine learning model in helpdesk ticketing systems.

Keywords: *Machine Learning, Helpdesk, Ticketing System, Literature Review*

1. INTRODUCTION

Helpdesk system has been used as a tool to provide users information about the system's structure, a web-based helpdesk offers IT Management Services from managing services, user role management, ticketing services and reporting services. [1]. Helpdesk system has been proven to be the most preferable tools for customer services, placing at the first place above FAQ and live chats, according to Hubspot state of service survey on 2020 [2]. This shows that the need for Helpdesk technology in customer service has the largest distribution, while stand-alone ticketing technology ranks lower in the survey. According to the survey, this is because modern helpdesk systems already include ticketing features. Data from [2] shows that the satisfaction of using the helpdesk has increased from 2019 until 2020 according to several factors such as increased productivity and it is easier to manage. A helpdesk with knowledge-based system

can be used as a good link to connect customer complaints with tickets and helpdesk Machine learning has been commonly used as a tool to improve help desk ticketing services by critically route ticket responsibility to corresponding ticket admin to be handled and minimizing resolution time [1], machine learning can also predict ticket resolution complexity of an issue that occur [3].

Many companies have also provided helpdesk system services such as Fresh Service helpdesk, Servicetonic Helpdesk, Manage Engine, Zoho helpdesk, Tribu.Ai, etc. Some of helpdesk vendor has been implementing machine learning in their services, for example tribu.ai has been implementing automatic ML prioritization, ticket classifier and predicting ticket responses based on ML, while ManageEngine ServiceDesk has been implementing issues predicting feature, improved search capabilities and issues classifier based on ML. These examples shows that ML has been implemented on supporting ticketing services in helpdesk systems

provided by those vendors by predicting, classifying and improving ticket solutions.

However, there is a possibility that machine learning could be used in several other elements in helpdesk ticketing systems. Core activities in helpdesk ticketing systems are ticket creation, ticket management, ticket reporting, ticket monitoring and ticket assignment [4]. The main objective in this paper is to review relevant works about machine learning use cases in helpdesk system and why further research is important to enhance helpdesk ticketing system by creating a function mapping of helpdesk ticketing core activities that has been research and what core activities has not been research with ML implementation. This paper aims to be one of the fundamental sources for further research about machine learning helpdesk use cases, therefore it is required to study various literature about machine learning presented in this paper. This paper will also present most effective methods to improve helpdesk ticketing services using machine learning algorithm and presenting common challenges and future works related to machine learning helpdesk ticketing systems by conducting a systematic literature review (SLR).

The research questions for this SLR are:

- RQ1. How machine learning could improve helpdesk ticketing systems function
- RQ2. What is the most frequently used machine learning algorithm models in helpdesk ticketing systems
- RQ3. What are the limitations of machine learning algorithms in helpdesk ticketing systems and further improvements that could be applied

This paper will show how ML implemented in various use cases of the study, identify what method of ML algorithm are used to solve the problems in previous studies and appraise degree of difficulty in each method [5] and comparing those findings with helpdesk ticketing core activities to show which areas have the most machine learning implemented on their process.

This paper will present data about how ML in previous studies is being used to solve their multiple purposes presented, this paper will present the algorithm used in those studies and the accuracy of the algorithm based on the years of the studies based on task category that has been defined. Data gathered in each study will be analysed using meta-analysis method to evaluate the evidence in each algorithm used in those reviewed studies [5] and then the summarization from individual studies will be

presented and discussed to answer this research question

This paper will also present data about limitations of each study sorted by year and shows how further studies about ML need to be researched to improve ML performance in helpdesk ticketing systems, these findings will help future works regarding machine learning implementation in helpdesk ticketing systems.

We hope this paper could provide systematical data of machine learning use cases and method in helpdesk ticketing function which could be useful for further research in other helpdesk ticketing function that has not been researched gathered from previous study and best practices to implement machine learning in helpdesk ticketing system.

This paper will be categorized into three sections: the first section is the introduction part, second section will be presenting methodology used by this paper and literature review presenting theoretical foundation to support the present research and third part will be discuss the results and conclusions from this study. This research paper is expected to provide an academic contribution to the field of literature systems and decision support in many sectors.

Helpdesk is part of a company that functions to provide documents, products, services or technology from the company. If it is connected to the application, it can be concluded that Helpdesk Ticketing Software is software used by the IT support team to develop, manage, and organize a list of customer problems automatically using a program [6].

A case study for helpdesk implementation at Schneider Electric company proved to have an efficient impact. The implemented web-based software can provide an easy solution to organize all documents related to each related department. This proposed feature can improve user satisfaction survey results. Before the project started, the satisfaction level about Help Desk access and service quality was 47% and became 97% in the second period of 2015 [7].

Machine Learning (ML) is an algorithms and statistical models that computer systems use to perform a specific task without being explicitly programmed by developers. Various applications that we use on a daily basis use learning algorithm. When a web search engine, such as Google, is used to search the internet, one of the reasons it works so well is due to a learning algorithm that has learned how to rank web pages. These algorithms are used for a variety of purposes, including data mining, image processing, predictive analytics, and so on [8].

Helpdesk with machine learning allows users to report issues, submit service requests, and communicate with IT agents through help desk tickets. It can also support prioritizing tickets, classifying tickets, assigning tickets to agents, editing ticket descriptions using a rich text editor, exchanging data for ticket description among participators, and changing ticket status. In addition, it allows managing services, managing user roles, generating automated email notifications, and generating reports that assist management in decision-making. Machine Learning-Based Ticket Classification is the linking of tickets with services or sub-services quickly with a minimum of manual human intervention, so as to reduce processing time, while the wrong ticket service classification will slow down the service process. [9].

Machine Learning can be used to improve the ticket resolution process by identifying tickets that require further input for resolution. Role-Based Models can then be used to identify possible inputs for these tickets. This work proposes a similar strategy. Using machine learning, this study identifies tickets with difficult resolutions. [10].

Machine learning needs algorithm to retrieve, modify and indexed, combining and run various function in Machine Learning. Supervised Learning is a learning function for Machine Learning that maps inputs to outputs based on a given example of input-output pairs. Supervised Learning is a function of the training data labels that exist in a set of training examples. This algorithm requires external assistance from a supervisor. The input dataset will be separated into two parts: train and test. There will be an output variable in the training dataset that must be predicted or categorized. draws a pattern from the training dataset and applies it to the test dataset for prediction or classification. [11]

Unsupervised learning are methods to be used when the incoming data does not have any labels to go with it. In such cases, a method known as clustering is used to uncover some relevant structure in the incoming data. So, after we find this structure in the form of clusters, groups, or other fascinating subsets, this composition will be employed for tasks such as providing a helpful summary of the input data and possibly displaying the structure. Unsupervised learning enables us to tackle issues with little or no knowledge of the end result.[11]

Machine Learning can be used for prioritization purposes, in this case the researchers tested 4 Machine Learning algorithm methods to carry out prioritization such as: support vector

machines (SVM), gradient boosting decision trees (GBDT), artificial neural networks (ANN) and Lambda Rank with neural networks (LRN). And in this study resulted that LRN has the best performance in conducting prioritization analysis speed and accuracy in conducting prioritization followed by SVM in second place and ANN in third place. [12]

The role of machine learning algorithms in improving ticketing services in helpdesk, according to recent studies, there are 6 machine learning algorithms that has been used in helpdesk ticketing system such as: KNN, Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes, Random Forrest (RF) and Deep Neural Network (DNN), coupled with unsupervised methods and standard algorithms using manual rule classification, and in this study shows that SVM and RF algorithms are the algorithms that have the best performance among others, it is also shown that the performance of applying machine learning has a better performance compared to manual classification algorithms. [13]

2. METHOD

In order to understand the state-of-the-art of the literature comprehensively, we will be using three stages procedure of literature review by [14]: planning the literature review, conduct the literature review and report the result of the literature review. To plan the systematic literature review, we will be using modified PSALSAR Framework [15] as follow

2.1 Protocol Search

Data will be collected from various journal data from the Scopus database and then filtered using Prisma flowchart in figure 1.

Records for identification will be searched from Scopus database starting from 2006-2022 with search terms as follow:

Search terms in Scopus		
Field Tag	Tittle, Abstract and Keywords	TITLE-ABS-KEY (machine AND learning AND (helpdesk OR ticketing))
Document Type	Conference Paper, Article, Conference Review	

Language	Limit to English	
Subject Area	Limit to Computer Science	
Full Search Boolean	TITLE-ABS-KEY (machine AND learning AND (helpdesk OR ticketing)) AND PUBYEAR > 2008 AND PUBYEAR < 2023 AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cr") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "re")) AND (LIMIT-TO (LANGUAGE , "English"))	

To analyze the data, we will be using a SLR approach using data analysis method based on PSALSAR Framework such as quantitative, descriptive and narrative analysis to answer all of research question defined in this paper. Data will be assessed manually by extracting necessary data manually from each individual study

2.4 Report

This paper will report the result by building the narrative conclusion based on findings and analysis and summarizing the report result of further studies

3. RESULT AND DISCUSSION

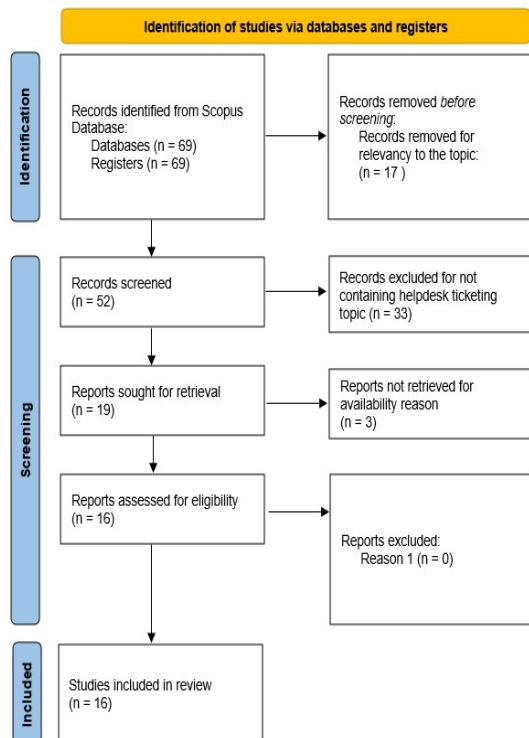
RQ 1: How machine learning (ML) could improve helpdesk ticketing system function

To answer this RQ all studies will be reviewed and classified into 3 main categories based on each machine learning algorithm function to solve its problems: Classification is a function of Machine Learning to automatically categorize a vast number of objects based on specific designated field [16], Regression is a set of technique of Machine Learning to estimate the relationship of an objects according to related variables [17] and Clustering is a simpler technique of Machine Learning to categorize an object without datasets based on the scenario, clustering methods can only separate objects based on the similarities and differences in the properties of the objects [18].

The distribution of 16 papers by publication year since 2008 until 2022 shows that classification category has the largest frequency, almost of the study use classification algorithm to perform various tasks from ticket identification, predict ticket, prioritize ticket and ticket alert using classification method, regression function has not been used in reviewed study and clustering has been used once for alerting error.

This difference in numbers could occur because the function of the tickets in the helpdesk itself, ticket data already has structured attributes to differentiate one ticket to other, therefore classification method is ideal and suitable for various tasks and use cases in ticketing helpdesk system using machine learning, the most common challenges for classification methods are the size of the sample data depending on cases for each study, for example in 2018 [19] found although they have tested 1.000.000 sample data, it need to be tested on much larger scale to achieve more accurate result and [20] found that legacy data was important and could hinder the ML model learning if it were not properly managed beforehand. But early on 2008

2.2 Appraisal and Synthesis



Document appraisal and synthesis will be conducted using Prisma flowchart consisting of 3 steps: Paper identification, screening and final paper included for this study

Figure 1. PRISMA Flowchart for Systematic Literature Review

2.3 Analysis

[21] classification ML model for ticketing have not been able to complete and only resulting on work in progress with positive results.

On the other hand, classification could also be perform using a smaller amount of training data compared, in 2014 [22] create a ML model based on 10.000 tickets data and acquire a 92% accuracy using MNB algorithm and in 2022 [23] developed a ML model based on maximum 81301 tickets data with 82,5%-91,24% with Hadoop algorithm.

It is also worth to mention that clustering method also used in [24] to create alert for ticket error, clustering method above are being used as a supplementary method for error alert because classification method has also been used in that study aimed at determining the correlation between the other two given alarms, based on their frequency and time of arrivals, these other 2 alarms are being modelled based using classification method

Table 1 Distribution of ML Algorithm Function by Year

Year	Machine Learning Algorithm Function	
	Classification (CS)	Clustering (CL)
2008	[21]	
2013	[25]	
2014	[22]	
2015	[26]	
2017	[27]	
2018	[28], [29]	
2019	[30]	[24]
2021	[31] [32]	
2022	[33]	
Total	8	1

Furthermore, to enhance the understanding of each study, this paper will present each scope of each study presented in Table 2 below to provide more detailed information on each study. Table 3 will summarize scope of studies in [34]Table 2 based on each unique category. Information of Table 3 will show how ML solve the problems that wanted to be solved or new function to be achieved using each ML algorithm function.

Based on these findings, ticket identification has the greatest number of studies and solved using ML classification function in ML, these findings are aligned with data on Table 1 because ticket identification is the first step to implement ML algorithm in helpdesk ticketing systems, and then it will be further improved to perform more various task such as ticket prediction,

ticket prioritization, ticket alert for anomaly and ticket resolution predictions.

Table 2 Distribution of ML Object

Year	Reference	Scope of Studies
2008	[21]	Prioritize and Solve Tickets
2013	[25]	Ticket Identification
2014	[22]	Ticket Identification
2015	[26]	Ticket Identification
2017	[27]	Ticket Identification based on image
2018	[19]	Anomaly Detection
2018	[28]	Predict Ticket Content
2018	[20]	Auto alert for monitoring
2018	[29]	Error Resolve, Ticket Classification
2019	[24]	Anomaly Detection
2019	[30]	Ticket Identification
2021	[35]	Auto alert for monitoring
2021	[31]	Ticket Identification
2021	[32]	Predict Ticket Trouble
2022	[23]	Predict Ticket Resolution
2022	[33]	Ticket Identification by problem classification

Table 3 Reference Scope Category

Scope Category	ML Category	Reference	Ticket Core Activities
Ticket Alert	Classification (CS)	[19], [30], [35], [20]	Ticket Monitoring
Ticket Identification	Classification (CS)	[25], [22], [26], [27], [31], [33], [29]	Ticket Creation
Ticket Prediction	Classification (CS)	[28], [32], [21], [29], [23]	Ticket Management
Ticket Prioritization	Classification (CS)	[21]	Ticket Assignment, Ticket Management
Ticket Alert	Clustering (CL)	[24]	Ticket Monitoring, Ticket Management

Table 3 also shows that ticket management, ticket creation and ticket monitoring have the greatest number of studies compared to ticket assignment. This shows us that their ML features in ticket assignment activities need to be further researched and applied with other unique ML function to broaden helpdesk ticketing performances in ticket assignment.

RQ2: *What is the most frequently used machine learning algorithm models in helpdesk ticketing systems*

To answer this question this paper will perform a meta-analysis method by presenting summarization of best applied model on each study and then combined the result of each study to answer RQ2 [36]. Data extracted from each paper will be presented in Table 4, this table show the average accuracy from best model applied on studies between 2008-2022 about ML on helpdesk ticketing systems.

Table 4 Algorithm Model Meta-Analysis Between 2008-2022

Referen ce	Algorithm Model	Accurac y (%)	Categor y
[21]	Decision Tree C4.5	94,10%	CS
[21]	Naïve Bayes	97,10%	CS
[25]	Conditional Random Field	88,74%	CS
[22]	MNB	92.6%	CS
[26]	KNN	87,50%	CS
[27]	Genetic Algorithm Backpropagati on Neural Network (GABPNN)	97,18%	CS
[19]	KNN	97,81%	CS
[28]	KNN	99,70%	CS
[20]	Naïve Bayes	97,00%	CS
[20]	SVM	77,00%	CS

[29]	Bagging Decision Tree	92,00%	CS
[24]	MAYOR (Decision Tree Based)	88,00%	CL
[30]	SVM	85,00%	CS
[35]	Fuzzy Logic	92,29%	CS
[31]	2 Step Neural Network	86,70%	CS
[32]	SVM with TD-IDF	81,40%	CS
[23]	Naïve Bayes	91,24%	CS
[33]	Decision Tree	91,00%	CS

Table 4 shows that Decision Tree is the most used algorithm from studies between 2008-2022 with accuracy range from 88%-94%, followed by Naïve Bayes 91%-97,10% and SVM 81%-85%. But Naïve Bayes shows the highest average accuracy followed by KNN 87%-97,18%. Decision Tree has the highest accuracy for classification purpose and it is worth to mention that MAYOR algorithm, a decision tree-based algorithm could be used for clustering purpose with high accuracy.

On the side note, MNB is uniquely implemented because it is compatible with XSEDE environment, Extreme Science and Engineering Discovery Environment (XSEDE) is an integrating framework system for advanced digital resources, to make it easier to use and help more people to use them which makes it easier for researchers to gain access to the services they need and move their work across different application in the same environment[37], this study is unique because it shows how XSEDE contribute in helpdesk ticketing system and what algorithm best suited for the task.

Fuzzy logic is a unique algorithm to be implemented because it is best suited for NLP data processing, and neural network, a deep learning method have been used in helpdesk ticketing system with high percentage of accuracy.

RQ3: *What are the limitations of machine learning algorithms in helpdesk ticketing systems*

To answer this research question, data from all studies in this paper will be presented with 2 sections: current limitations and further improvement, but since not all studies conclude

their limitations or further improvements, this systematic literature review will narratively discuss and analyze critical points in the research or the cause of limitations and further improvements in some studies that need to be concerned and could be done in further research.

Table 5 Limitations of Studies

Year	Reference	Limitations
2008	[21]	Noisy information in the dataset causing the search for behavior patterns in tickets to find those that are not caused by an isolated problem
2014	[22]	The Porter stemmer model could not deal with subtle technical terms, more specific keyword could be developed in the dataset to increase accuracy
2015	[26]	Due to algorithm restrictions, binary classification of tickets could be extended into a ticket ranking system that will allow the most important tickets to be analyzed first
2017	[27]	Need more powerful CPU to shorten learning time
2018	[28]	Periodically edit the training set to maintain acceptable prediction times is needed
2018	[20]	Missing ground truth annotation of the dataset for the sentiment analysis part
2018	[29]	Dataset limitations
2019	[24]	Evaluate the system's prediction capability across longer time scales (such as multiple days), as well as the effect of changing system settings on the system's accuracy or computation tractability
2019	[30]	Handling the initial unstructured noisy data

2021	[35]	There are numerous noisy data created by user sickness from user voice input signals and launch this e-governance application integrating all governmental sectors benefiting human life quality and connecting government individuals closely with the public
2021	[31]	The database requires an updated table incorporating the additional categories and subcategories to provide employees and clients with a wide range of options from which to choose and express their problem in a more meaningful manner. The database requires an updated table incorporating the additional categories and subcategories to provide employees and clients with a wide range of options from which to choose and express their problem in a more meaningful manner
2021	[32]	Data accuracy could be improved from training dataset quality
2022	[23]	Standard Operating Procedures (SOP) and service quality in a specific service zone need to be reviewed to produce more accurate data sets for machine learning
2022	[33]	Dataset size too small

Table 6 Further Improvements of Studies

Year	Reference	Further Improvements
2008	[21]	Solving noisy information problem will allow to drastically decrease the number of tickets that are active at each moment, locating those tickets that refer

		to the root problem (those higher in the hierarchy) and whose resolution will allow the automatic closing of all those that descend from it			components and link the investigated system with the existing ticketing system is needed. Using unsupervised machine learning to leverage clustering techniques like centroid or density-based methods, as well as other topic models like LDA or TKM, to further refine ticket categorizations over time and perform finer granular sentiment analysis stating sentiments per topic or ticket.	
2014	[22]	To expand the asset configuration database and connect all server-related information. configuration database to determine whether a specific combination of configurations is prone to failure by Obtaining a complete perspective of information from diverse business units is critical to the success of service delivery		[30]	Deep learning-based classification models to automatically classify the service desk tickets and to investigate its performance on our IT infrastructure ticket data and larger dataset training	
2017	[27]	Explore the combination of standard machine learning classifiers with manually created keyword lists in order to further increase the classification accuracy. In addition, the full integration of automatic classification within business processes will allow for a continuous annotation of support tickets		[31]	Explore alternate classification techniques like ensemble of classifiers to further improve the performance of the model and exploring more machine learning techniques required to handle data related challenges and further focus on larger data	
2018	[28]	More various studies are required to establish which BPNN variables or combinations of those variables should be improved, as well as what methods and operations should be used in the GA process to achieve better and optimal results		2021		
2018	[20]	Further evaluation of the k-NN model and further testing of prototype model		2022	[33]	Collect more dataset and improve stop-gap model
2018	[29]	Testing of the recommended different approach and implement new solution prototype				
2019	[24]	To manage tickets in an automated manner testing of the entire cycle for other				

Table 7 Summary of Study Limitations

Year	Reference	Limitation Category	Details
2008	[21]	Dataset Problem	Unstructured Data
2014	[22]	ML Model Limitations	Model need very specific link to be defined
2015	[26]	ML Model Limitations	Classifiers cannot categorize each individual

			ticket accurately
2017	[27]	ML Model Limitations	Slower feedforward time due to the number of hidden neurons
2018	[28]	ML Model Limitations	Regular pruning and editing in the model to sustain long-term usage
2018	[20]	Dataset Problem	Missing Values
2018	[29]	Dataset Problem	Missing Values
2019	[24]	ML Model Limitations	Time Consuming
2019	[30]	Dataset Problem	Unstructured Data
2021	[35]	Dataset Problem	Missing Values
2021	[31]	Dataset Problem	Missing Values
2021	[32]	Dataset Problem	Missing Values
2022	[23]	Dataset Problem	Unit Values Transformation
2022	[33]	Dataset Problem	Other errors, dataset size too small

Table 7 consist of the summarization of table 6 and it is classified into 2 categories, dataset problem and ML model limitations. ML model limitations briefly described in the table while for dataset problems is categorized based on real life datasets problem category based on [38], there are unstructured data, missing values, unit values transformation problems and other problems such as data quantity problem that exist in the summary of the study.

Although there are some limitations on ML Model that has been developed in each study, all studies shown positive results that has been already discussed in RQ I, but the case of limitations need to be reviewed individually on each study.

Table 7 show us that it is critical to maintain a good dataset and remove noisy data as much as possible before applying it into practice datasets for machine learning training, because noisy and unstructured data could potentially affect

machine learning accuracy and its relevancy to run its purposes.

It is also crucial to run a thorough research on applied machine learning algorithm that will be used because all machine learning algorithm has its specific excellency and limitations that need to be concerned before applying it, this information could be researched by looking at relevant previous study.

4. CONCLUSION

Based on the systematic literature review related to machine learning approach for helpdesk ticketing system. There are various scientific methods and algorithm to improve helpdesk ticketing system performance using machine learning based on varied purposes designed specifically to solve different problem and cases.

This research focused on articles from Scopus Indexed article from different sources such as: Elsevier, Research Gate, and Semantic Scholar online databases for articles published in the English language from 2006-2022 with these Boolean keywords (Machine AND Learning AND (Helpdesk OR Ticketing)). This research presenting unique information analyzed by PSALSAR Framework using quantitative, descriptive and narrative analysis to answer all of research question defined in this paper performed manually by studying each paper will present more in-depth information regarding related papers analyzed in this systematic literature review.

It concluded from RQ1 that Classification is the most used machine learning function and it can be applied in any form of helpdesk to classify ticket based on its similarities for automatic ticket identification. More specific use cases are needed to perform machine learning regression and clustering purposes because various task of the use cases can be accomplished by using classification method in helpdesk ticketing system because of the properties of the ticket itself, while clustering function in helpdesk ticketing system could be used to create an auto alert for ticket trouble and ticket prioritization that does not need any training dataset for unsupervised machine learning method. Based on the number of research in RQ1, regression and clustering function using machine learning in helpdesk ticketing need to be further research to explore more effective design and implementation technique. Based on core activities defined with machine learning implementation on each activity, it can be concluded that machine learning based ticket assignment is the least activities applied with

machine learning function, further research could focus on these activities in order to create a more sophisticated function in helpdesk ticketing systems which supports various user needs.

RQ2 shows how all applied machine learning methods in helpdesk ticketing system has a high accuracy in performing their respective function, all of the applied model has 88% accuracy minimum and 97% highest accuracy, based on this fact it can be concluded that more advanced machine learning use cases in helpdesk ticketing is ready to be researched in order to increase its efficiency in performing their tasks. It is worth to mention that deep learning method, the more advanced form of machine learning has also been applied and documented in 2 of the study in this literature review which resulted in very positive results based on the needs and resources needed. Future machine learning implementation in helpdesk ticketing systems could reflect on findings above to effectively classify the most suitable ML algorithm approach before proceeding to design and testing phase.

Database management also proven to be critical in RQ3, because machine learning has been emerging as an effective tool to enhance the business process, it is crucial for any company to properly managed their database to make it easier for machine learning development when needed. Determining machine learning model classifier has been proven to be one of the most critical steps to implement a machine learning model, because all of the problems in machine learning above started because there is a problem with the applied classifier. Further development also needed to test different machine learning model in other helpdesk ticketing sequences as this topic need as many research as possible to further enhance the machine learning performance in helpdesk ticketing systems.

While analyzing specific topic manually using quantitative, descriptive and narrative analysis based on PSALSAR framework to answer all of research question have its own advantages, this research is having limitations because of its limited sources for its very specific topic regarding machine learning in helpdesk system.

Another systematic literature review will be needed to review the possibility of a new technology based on machine learning such as AI and another use cases and function of machine learning that can be implemented for helpdesk machine learning ticketing systems since helpdesk vendor has been improving their services in helpdesk-ticketing systems by implementing new

technology such as smart dashboard an AI based chatbot.

REFERENCES:

- [1] F. Al-Hawari and H. Barham, "A machine learning based help desk system for IT service management," *Journal of King Saud University - Computer and Information Sciences*, Jul. 2019, doi: 10.1016/j.jksuci.2019.04.001.
- [2] Y. Chen, "The State of Customer Service in 2020 Table of Contents by Hubspot," 2020.
- [3] J. Schad, R. Sambasivan, and C. Woodward, "Predicting help desk ticket reassignments with graph convolutional networks," *Machine Learning with Applications*, vol. 7, p. 100237, Mar. 2022, doi: 10.1016/j.mlwa.2021.100237.
- [4] H. Yamaoka, K. Yamamoto, T. Nagai, and H. Masuda, "Case study of implementing an IT service desk ticketing system at small computer center," in *Proceedings ACM SIGUCCS User Services Conference*, Association for Computing Machinery, Oct. 2019, pp. 140–144. doi: 10.1145/3347709.3347820.
- [5] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J Bus Res*, vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.
- [6] Bob Wooten and Robert J. Wooten, *Building & Managing a World Class IT Help Desk*. Michigan: Osborne/McGraw-Hill, 2001.
- [7] S. Serbest, Y. Goksen, O. Dogan, and A. Tokdemir, "Design and Implementation of Help Desk System on the Effective Focus of Information System," *Procedia Economics and Finance*, vol. 33, pp. 461–467, 2015, doi: 10.1016/s2212-5671(15)01729-3.
- [8] B. Mahesh, "Machine Learning Algorithms-A Review," *International Journal of Science and Research*, 2018, doi: 10.21275/ART20203995.
- [9] F. Al-Hawari and H. Barham, "A machine learning based help desk system for IT service management," *Journal of King Saud University - Computer and Information Sciences*, Jul. 2019, doi: 10.1016/j.jksuci.2019.04.001.
- [10] O. Farooq and P. Gupta, "Machine Learning Approaches for IoT-Data Classification." [Online]. Available: <https://ssrn.com/abstract=3563391>

- [11] S. M. Basha and D. S. Rajput, "Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap," in *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, Elsevier, 2019, pp. 153–164. doi: 10.1016/B978-0-12-816718-2.00016-6.
- [12] D. Marijan, "Comparative Study of Machine Learning Test Case Prioritization for Continuous Integration Testing," Apr. 2022, [Online]. Available: <http://arxiv.org/abs/2204.10899>
- [13] S. Fuchs, C. Drieschner, and H. Wittges, *Improving Support Ticket Systems Using Machine Learning: A Literature Review*. [Online]. Available: <https://hdl.handle.net/10125/79570>
- [14] D. Tranfield, D. Denyer, and P. Smart, "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review *."
- [15] W. Mengist, T. Soromessa, and G. Legese, "Ecosystem services research in mountainous regions: A systematic literature review on current knowledge and research gaps," *Science of the Total Environment*, vol. 702. Elsevier B.V., Feb. 01, 2020. doi: 10.1016/j.scitotenv.2019.134581.
- [16] S. Chowdhury and M. P. Schoen, "Research Paper Classification using Supervised Machine Learning Techniques," in *2020 Intermountain Engineering, Technology and Computing, IETC 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/IETC47856.2020.9249211.
- [17] H. Saleh, J. A. Layouts, and S. A. Republic, "Machine Learning-Regression Design and Implementation of a Reaction Wheel-Actuated Cube View project Hybrid robots View project Machine Learning-Regression", doi: 10.13140/RG.2.2.35768.67842.
- [18] P. W. J. Glover, O. K. Mohammed-Sajed, C. Akyüz, P. Lorinczi, and R. Collier, "Clustering of facies in tight carbonates using machine learning," *Mar Pet Geol*, vol. 144, Oct. 2022, doi: 10.1016/j.marpetgeo.2022.105828.
- [19] B. Božić, A. Ríos, and S. J. Delany, "Comparing tagging suggestion models on discrete corpora," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2018, pp. 15–23. doi: 10.1145/3282373.3282386.
- [20] R. Qamili, S. Shabani, and J. Schneider, "An Intelligent Framework for Issue Ticketing System Based on Machine Learning," in *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOCW*, Institute of Electrical and Electronics Engineers Inc., Nov. 2018, pp. 79–86. doi: 10.1109/EDOCW.2018.00022.
- [21] Y. Temprado, F. J. Molinero, C. García, and J. Gómez, "Knowledge discovery from trouble ticketing reports in a large telecommunication company," in *2008 International Conference on Computational Intelligence for Modelling Control and Automation, CIMCA 2008*, 2008, pp. 37–42. doi: 10.1109/CIMCA.2008.116.
- [22] G. Son, V. Hazlewood, and G. D. Peterson, "On automating XSEDE user ticket classification," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2014. doi: 10.1145/2616498.2616549.
- [23] F. Che Yayah, K. Imran Ghauth, and C.-Y. Ting, "The Performance of Classification Method in Telco Customer Trouble Ticket Dataset."
- [24] Institute of Electrical and Electronics Engineers, *ICCCN 2019: the 28th International Conference on Computer Communication and Networks: July 29 - August 1, 2019, Meliá Valencia Hotel, Avenida de las Cortes Valencianas, 52, Valencia, Spain*.
- [25] C. S. Hong *et al.*, *Proceedings of the 2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013): 27-31 May 2013, Ghent, Belgium*.
- [26] Zinner T.; Lemmerich F.; Schwarzmann S.; Hirth M.; Karg P.; Hotho A., *Big Data Analytics and Knowledge Discovery*, vol. 9263. in *Lecture Notes in Computer Science*, vol. 9263. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-22729-0.
- [27] J. Tarigan, Nadia, R. Diedan, and Y. Suryana, "Plate Recognition Using Backpropagation Neural Network and Genetic Algorithm," in *Procedia Computer Science*, Elsevier B.V., 2017, pp. 365–372. doi: 10.1016/j.procs.2017.10.068.

- [28] B. Božić, A. Ríos, and S. J. Delany, “Validation of tagging suggestion models for a hotel ticketing corpus,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Nov. 2018, pp. 15–23. doi: 10.1145/3282373.3282386.
- [29] Paramesh S.P; Ramya C; Shreedhara K.S, *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*. IEEE, 2018.
- [30] S. P. Paramesh and K. S. Shreedhara, “Automated IT service desk systems using machine learning techniques,” in *Lecture Notes in Networks and Systems*, Springer, 2019, pp. 331–346. doi: 10.1007/978-981-13-2514-4_28.
- [31] F. Islam and D. Logofătu, “Comparative Study of Machine Learning Algorithms for Performant Text Analysis in a Real World System,” in *Communications in Computer and Information Science*, Springer Science and Business Media Deutschland GmbH, 2021, pp. 162–173. doi: 10.1007/978-981-16-1685-3_14.
- [32] F. Al-Hawari and H. Barham, “A machine learning based help desk system for IT service management,” *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 702–718, Jul. 2021, doi: 10.1016/j.jksuci.2019.04.001.
- [33] R. Nicholls, R. Fellows, S. Battle, and H. Ihshaish, “Problem Classification for Tailored Helpdesk Auto-replies,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 445–454. doi: 10.1007/978-3-031-15937-4_37.
- [34] S. Serbest, Y. Goksen, O. Dogan, and A. Tokdemir, “Design and Implementation of Help Desk System on the Effective Focus of Information System,” *Procedia Economics and Finance*, vol. 33, pp. 461–467, 2015, doi: 10.1016/s2212-5671(15)01729-3.
- [35] S. Leelavathy and M. Nithya, “Public opinion mining using natural language processing technique for improvisation towards smart city,” *Int J Speech Technol*, vol. 24, no. 3, pp. 561–569, Sep. 2021, doi: 10.1007/s10772-020-09766-z.
- [36] C. Hansen, H. Steinmetz, and J. Block, “How to conduct a meta-analysis in eight steps: a practical guide,” *Management Review Quarterly*, vol. 72, no. 1. Springer Science and Business Media Deutschland GmbH, Feb. 01, 2022. doi: 10.1007/s11301-021-00247-4.
- [37] F. Bachmann *et al.*, “XSEDE Architecture Overview A Level 1 and 2 Decomposition for Software Developers and Integrators CORE View metadata, citation and similar papers at core.ac.uk provided by Illinois Digital Environment for Access to Learning and Scholarship Repository,” 2014.
- [38] S. H. Eбенуwa, M. S. Sharif, M. Alazab, and A. Al-Nemrat, “Variance Ranking Attributes Selection Techniques for Binary Classification Problem in Imbalance Data,” *IEEE Access*, vol. 7, pp. 24649–24666, 2019, doi: 10.1109/ACCESS.2019.2899578.