International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Recommending Automation Solutions for Information Technology Infrastructure Issues

Sandeep Bhutani

Shri Krishnan Institute of Engg, & Technology, Kurukshetra University, Haryana – 136119

A B S T R A C T

The proposed system aims to improve the efficiency and cost-effectiveness of resolving technology infrastructure issues by using advanced techniques such as machine learning and natural language processing to automatically identify issues that can be solved through automation and predict solutions for them. By eliminating the subjectivity and bias that come with manual assessments, this system can also predict potential savings from automation and the time it will take to reach the break-even point, allowing for selective application of automation to issues that provide the most benefits in the shortest period of time.

The use of automation solutions in a targeted and selective way can enhance the overall performance of technology infrastructure operations and make sure that human resources are utilized for issues that are more intricate and/or cost-effective to solve manually.

A system is needed that can sort technology infrastructure problems and pinpoint matching automation answers already linked to a specific group/cluster/category. Moreover, by taking into account both the context and verbal meaning of the text from the infrastructure data when categorizing infrastructure issues, the disclosed systems and methods may boost the precision of classification. This may lower the expenses associated with utilizing incorrect automation solutions to fix infrastructure issues.

Keywords: IT Infrastructure automation, Machine Learning, Artificial intelligence, Automation, Ticket prediction

1. Introduction

This issue we are going to discuss is related to technology infrastructure issues and its solutions. More specifically, it is related to a system for prediction of automation solutions for infrastructure issues in information technology field.

The IT infrastructure has been developed over a long period of time and is considered old. Many companies use support ticketing systems to identify problems in their technology infrastructure. These systems are filled out by users, such as employees or customers, who list issues or incidents that need to be resolved. As the company and user base expands, the number of issues to be resolved may also increase. To improve the efficiency and/or reduce the cost of solving these issues, companies try to automate some of the solutions. Traditional methods for identifying automated solutions involve a manual review of existing support tickets to determine which ones can be automated. However, this approach can be time-consuming, taking several days or weeks, depending on the number of support tickets that need to be reviewed. Additionally, manual assessments may be biased and have difficulty in producing standardized results and quantifying the outcome.

There is a need for a system that can addresses and solve the above discussed shortcomings.

2. Literature Review

Information technology infrastructure components are part of almost every industry and so this problem occurs in many industries, domains and businesses. Multiple people have attempted multiple times, to solve this problem. Following research papers and patents have been published which try to address and solve this issue, similar issue and some part of the issue in some ways.

Paramesh, S. P., in his paper “A Deep Learning based IT Service Desk ticket classifier using CNN” published in 2022 [1], Assigning tickets to the proper resolver group is a critical task in IT Service management. Manual assignment can lead to wrong assignations, reassignments, delays, and interruptions. This study proposes using a deep neural network with Convolutional Neural Networks (CNN) to automatically classify service desk tickets. The CNN model identifies important features in ticket descriptions using word embeddings, resulting in accurate category predictions. Testing with real-world data shows that the proposed model outperforms traditional models like SVM, NB, Logistic Regression, and KNN. Implementing this deep learning method offers benefits such as efficient ticket assignment, faster resolution, increased productivity, improved customer satisfaction, and uninterrupted business operations.

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: author@institute.xxx

Ali Zaidi, SS, in paper "A multiapproach generalized framework for automated solution suggestion of support tickets" [2], Customer support departments play a crucial role in a company's success and reputation. However, handling a large volume of customer tickets with limited support staff can be challenging. To address this, this study presents a framework consisting of four components: data preprocessing, actions extractor, resolution predictor, and evaluation. The actions extractor module detects actionable phrases from resolution text, while the resolution predictor offers two different pipelines: a similarity search model and an end-to-end model. The end-to-end model outperforms the similarity search-based techniques, accurately predicting actions for resolution. Real-world IBM ticket data was used for analysis and experimentation. Implementing this framework can improve customer support by suggesting effective actions for new tickets, enhancing efficiency and performance.

Fauzy Che Yayah, in publication "The automated machine learning classification approach on Telco trouble ticket dataset" [3], This paper presents a solution for the telecommunications industry to address the increasing number of trouble tickets generated by digital technology. The focus is on classifying early code resolution for each ticket, improving service restoration time and customer satisfaction. The research explores traditional models and proposes modifications to enhance their accuracy. The Automated Predictive Engine (AutoPE) is introduced as a method that combines machine learning techniques and grid search optimization to achieve a significant accuracy increase, from 5% to 38%. Multiple classification modeling techniques are employed, such as Deep Learning, Random Forest, XGBoost, Gradient Boosting, and Extremely Randomized Trees. The AutoPE model surpasses the traditional model in accurately classifying early code resolution in the telco trouble ticket dataset. This solution can reduce costs and enhance overall operations for telecommunications companies.

In this patent "System and method for classifying and resolving software production incident" [4], Ryali proposes techniques for categorising and resolving software incident tickets created in production environments, it extracts a set of similar keywords from the incident ticket, and generates a query vector with respect to the incident ticket based on the synonyms or similar keywords comprise a system and method for categorising and resolving software production incident tickets. It compares the query vector and similar vectors generated from similar previous incident tickets, which is then further categorized the incident ticket into at least one of a positive automation incident ticket and a negative automation incident ticket. The vectors are derived from a set of keywords and the occurrences of those keywords in a set of historical incident tickets.

Ramkumar Balasubramanian and Sandeep Bhutani, in patent "System and method for recommending automation solutions for technology infrastructure issues" [5], suggested a system and set of techniques for intelligently developing automation plans for technology infrastructure operations. The system and the techniques comprise forecasting automated solutions and evaluating infrastructure issue data from support requests. The automation solutions are then subjected to a cost-benefit analysis. The cost-benefit analysis may be used to rank and suggest solutions.

In this solution, first the ticket data was preprocessed using NLP (Natural Language processing) techniques. Then PCA (Principal component analysis) is carried out to generate the orthogonal vectors. The PCA components are used for clustering. Here Louvain clustering unsupervised clustering technique was used, which is a network-based community detection technique. This technique produced better clustering than K-Means and also the number of clusters were also not needed to be pre-provided to the algorithm. The generated clusters were then passed to clustering algorithms. This classification had resulted in the ticket category which then could be mapped to already curated knowledge base.

There are pre-build automation solutions mapped to the ticket category in knowledge base. Which were used to calculate the cost-benefit analysis based on potential benefit that is provided by the automation artifact mapped in knowledge base.

In this paper "System and method for classifying and resolving software production incident" [6], Ryali [4] proposes techniques for categorizing and resolving software incident tickets in production environments. The approach involves extracting similar keywords from the ticket and generating a query vector based on these keywords. This vector is compared to vectors from previous tickets to categorize the current ticket as a positive or negative automation incident. The vectors are derived from historical incident tickets and the occurrence of keywords within them. This method improves upon Ryali's previous work by providing a system and method for effectively categorizing and resolving software production incident tickets based on keyword analysis and vector comparison.

Sabharwal et al., "System and method for assisting user to resolve a hardware issue and a software issue" [7], The paper relates to techniques that help a user troubleshoot and resolve hardware and software problems. The system will identify a specific cluster which can be related to the new ticket that is now received from the user, from a group of clusters. It will then recommend runbook scripts(s), from a repository of runbooks, that are relevant to the new ticket. His technique also identifies a new runbook script, which can be related to the new ticket received, from a set of repositories which are external to this recommendation system. It will then execute at least one of the runbook script(s) recommended by this engine or the new runbook script, related with the new ticket. After execution, the system automatically can create a document or log based on the execution of the new runbook script or one or more runbook script(s), which will assist user in resolving the issue he/she is working on.

Rathod et al. in "Classifying and routing enterprise incident tickets" [8], the disclosed system classifies and directs incident tickets by analyzing the related words within them. Categories of incidents are created based on word associations found in the tickets. Match scores are calculated by examining the presence of related words in each ticket, and based on these scores, the system assigns the ticket to a specific incident category. It then generates output to route the ticket within an incident management system based on its assigned category.

Srivastava, Krupa, in his paper "Automated ticket resolution" [9], The device stores instructions in memory and executes them using one or more processors. It communicates with a server to retrieve historical ticket data and generates data models for different ticket types. It also interacts with a client device to obtain ticket data for a project-related issue. Using natural language processing and a specific data model, the device classifies the ticket data and trains the model using machine learning techniques. It processes a subset of historical ticket data to generate recommended resolutions for the current issue. The device selects the most effective resolution and takes actions to implement it, including code generation, modification, or removal. Additionally, the device predicts and generates data for potential issues using machine learning and provides recommended resolutions for them.

Guven et al., in paper titled “Automated change monitoring and improvement recommendation system for incident reduction in information technology infrastructure” [10], The process collects information from a service management database, specifically change and incident tickets related to IT infrastructure. It identifies relationships between these tickets to determine factors affecting changes made to the infrastructure. Using these factors, the process generates recommendations for future changes. The recommendations are implemented and monitored, and if needed, adjusted based on the outcomes.

ALEKSANDRA REVINA, “IT Ticket Classification: The Simpler, the Better” [11], This study explores designing an IT ticket classification pipeline for tracking and resolving technical issues. It compares the performance of different text representation techniques (TF-IDF and linguistic features) and classification algorithms (decision trees, KNN, logistic regression, naive Bayes, and support vector machines) to predict the complexity level of an IT ticket (low, medium, or high). The findings indicate that linguistic features are more effective for representing IT ticket text, and feature selection is crucial for accurate predictions. The study also shows that simple algorithms can yield good results when paired with suitable linguistic features.

YANG, Libo in paper “Fuzzy Output Support Vector Machine Based Incident Ticket Classification” [12], presented method that aims to classify tickets (specifically incident tickets) to enhance maintenance efficiency and cost reduction. It utilizes fuzzy output support vector machine (FOSVM) for both multi-class and binary classification. By examining regions that are challenging for multi-class SVMs, the approach achieves more precise and dependable outcomes. Evaluation of real-world ticket data and benchmark data demonstrates the effectiveness of this technique.

Maksai, Andrii, in paper “Hierarchical Incident Ticket Classification with Minimal Supervision” [13], a new method for categorizing incident tickets is proposed to minimize the amount of manual labeling required while still producing accurate predictions. The approach is a two-step process that is using a combo of topic modeling and graph clustering in the step one, followed by either an active learning technique or another step of hierarchical clustering in step two. The effectiveness of the method is evaluated using real-world data sets and it is shown that traditional text classification methods are not appropriate for incident ticket texts.

Pradhan, Dr in “ITSM Using AI Chat-Bot and Data Visualizers” [14], This project enhances IT Service Management (ITSM) systems by integrating chatbots, data visualization, and a quality ticket system based on the ITIL framework. The system follows a three-step process: reporting, managing, and resolving issues, with the goal of restoring normal business operations. The incorporation of a chatbot enables users to address minor issues swiftly without creating a ticket. Data visualization facilitates ticket tracking and monitoring for management teams and users. The project also explores the application of natural language processing.

2.1 Inferences Drawn from Literature Review

As we can observe from various papers published in the area to solve these problems, following are most common shortcomings in existing solutions:

1. Some do not focus on specific to IT infrastructure issues
2. some takes change request or service request only in focus
3. most of these do not identify and execute the automation
4. Most of these are not using state of art semantic matching techniques
5. Does not target infrastructure specific issues

3. Problem Formulation and Proposed Work

The system described in this passage is designed to identify and predict automation solutions for technology infrastructure issues. By utilizing natural language processing (NLP) and machine learning (ML), it can analyze and classify issues, reducing subjectivity and bias in the assessment process. The system's ability to predict potential savings and estimate the break-even time helps in selecting the most cost-effective automation solutions, thus lowering technology infrastructure costs. By targeting specific issues for automation, the efficiency of technology infrastructure operations can be improved, and resources can be allocated to problems that require more complex or cost-effective manual solutions.

The system's features include analyzing information, extracting data features, predicting automation solutions, estimating cost savings, and ranking the solutions based on their cost-effectiveness. It leverages both traditional and advanced methods and incorporates industry-standard calculations and formulas to determine cost savings. It matches technology infrastructure issues with pre-existing automation solutions, considering the context and language of the issue data for improved accuracy. The system is flexible, allowing for variations and combinations of its components, and aims to enhance technology infrastructure operations while reducing costs through effective automation solutions.

Table 1 – Sample Tickets.

Sample Ticket Body	Activity	Industry	Automation Possible
Hello, My OS drive is full. I do not store any content on C drive, but still it is full. So my system is slow and I get disk full alerts also. Please fix	Disk Utilization	Generic	Yes
10.231.0.1!›3 CURRENT CPUUTILIZATION: 100PERCENT CONFIGURE:D CRITICAL CHANGED FROM GOOD TO ERROR	CPU Utilization (Auto Generated)	Generic, Oil & Gas	Yes
THEE THRESHOLD FOR THE PROCESSOR INFORMATION\PERCENT PROCESSOR TIME\ TOTAL PERFORh1ANCE COUNTER HAS BEEN EXCEEDED. THE VALUES "FHAT EXCEEDED THE THRESHOLD ARE: PER	CPU Utilization (Auto Generated)	Generic, Hospitality	Yes

4. Methodology

The automation recommendation system performs six tasks: analyzing IT service tickets, assessing automation feasibility, suggesting automation assets, calculating manual effort, providing a cost-benefit analysis, and predicting break-even time. It extracts and analyzes infrastructure issue data, predicting automation solutions for automatable issues and indicating when no solution is possible. The system classifies issues and determines if they can be resolved through automation, referring to a knowledge base for solutions. It conducts cost-benefit analyses by calculating expenses and projected savings, estimating manual effort without automation. IT incident data is uploaded via the data upload module, including issue details, dates, configuration item, and relevant information.

Once uploaded, data undergoes normalization and categorization based on group details. The automation use case classifier module sorts issue data into groups, using machine learning algorithms to classify it. Predictions are written to the output file. The system relies on an infrastructure automation knowledge base to determine if a support ticket can be automated, providing associated solutions. If the analysis falls below a threshold, an abstraction process allows manual review and labeling. The technology prediction module determines the optimal technology for each automatable issue, prioritizing root causes to avoid duplication. An automation catalog lists existing automations created using various tools like RPA and scripting technologies.

The automation strategy module evaluates data for automation based on factors such as time savings, activity intrusiveness, complexity, MTTR savings, downtime savings, ROI calculations, and SLA benefits. Preprocessing involves operations like lowercase conversion, tokenization, stop word removal, stemming, and lemmatization. The feature extractor generates extracted features through transformations and utilizes the automation knowledge base and linguistic validator. The classification engine creates a prediction model using machine learning techniques. The hetero sampler module provides stratified samples for abstraction and technology prediction. The cost analysis module estimates automation costs. The system categorizes issues into tasks, processes, and cognitive solutions. It prevents redundancy and extracts additional features using vectorization and tf-idf analysis. The feature extractor creates a vector space model using word dictionaries and Bert vectors, assessing similarity using Euclidean or cosine distances.

The Classifier predicts automation solutions using input features from modules like the Feature Extractor, including issue sequence, n-grams, distances, and similarity. It utilizes machine learning algorithms such as deep learning, Bert analysis, and ChatGPT. The Classifier categorizes issues into classes within the Vector model and associates them with known automation solutions. It can estimate costs and savings using techniques like complexity and occurrence-based cost analysis. The recommendation system prioritizes solution groups based on cost and savings metrics, suggesting solutions with high savings potential. Common machine learning frameworks like Scikit-learn and Tensorflow can implement these algorithms (e.g., BERT, K-Means, Louvain clustering, tf-idf).

To find out the break even analysis, solution needs a base data in terms of FTE (Full time employee) cost in USD. Following is the sample of data to be supplied to solution:

	Per hour	Per Month	Per Year
FTE Cost	$60	$10,560	$126,720

Basis the above cost of a human expert who can solve the problem manually and the cost of automation which is part of automation catalog, system does a pareto analysis to find out which processes should be automated and how much time it would take to realize the benefits of automation compared to the humans solving the tickets in absence of automation. Such results are automatically produced by solution, results of which are depicted in results section below.

5. RESULTS AND DISCUSSION

5.1 Datasets used

There are two possible sources of dataset - Private and public. Public dataset from kaggle has been used to train and test the model. Private dataset is held by organizations which contains sensitive information about their assets, so it is not easily possible to get access to that dataset.

From public, two main sources of data were:

1. https://www.kaggle.com/datasets/nikolagreb/small-itsm-dataset [16]

2.Synthetic data – Synthetic data was generated using ChatGPT

5.2 Results

System was able to correctly classify results on test set with following accuracy:

Technique	Accuracy
BERT based classification	92.3021%

With the help of automated pareto analysis system can tell which processes should be automated to realize a break even as shown below

System also suggests:

Following processes consume 80% of total manual cost: ['Process7', 'Process2', 'Process4', 'Process5']

And with same automated system indicates the time period in which the cost of automation will overcome the manual effort spend if automation is not carried out. Following illustration depicts the same:

References

1. Paramesh, S. P., and K. S. Shreedhara. "A Deep Learning Based It Service Desk Ticket Classifier Using Cnn."; October 2022
2. Ali Zaidi, SS, Fraz, MM, Shahzad, M, Khan, S.; “A multiapproach generalized framework for automated solution suggestion of support tickets”; Int J Intell Syst. 2022;
3. Fauzy Che Yayah*, Khairil Imran Ghauth, Choo-Yee Ting; “The automated machine learning classification approach on Telco trouble ticket dataset”; October 2021
4. Ryali et al.; System and method for classifying and resolving software production incident; US10067760B2; 2018
5. Ramkumar Balasubramanian; Sandeep Bhutani ; System and method for recommending automation solutions for technology infrastructure issues; US10904072B2; 2019
6. Ryali et al.; System and method for classifying and resolving software production incident; US20180032330A9; 2016
7. Sabharwal et al.; System and method for assisting user to resolve a hardware issue and a software issue; US10769043B2; 2018
8. Rathod et al.; US20200202302A1; Classifying and routing enterprise incident tickets; 2018
9. Srivastava, Krupa; Automated ticket resolution; AU2019202728B2; 2016
10. Guven et al.; Automated change monitoring and improvement recommendation system for incident reduction in information technology infrastructure; US10547507B2; 2015
11. Aleksandra Revina; IT Ticket Classification: The Simpler, the Better; Jan 2020
12. Yang, Libo. (2021). Fuzzy Output Support Vector Machine Based Incident Ticket Classification. IEICE Transactions on Information and Systems. E104.D. 146-151.10.1587 / transinf. 2020
13. Maksai, Andrii & Bogojeska, Jasmina & Wiesmann, Dorothea. (2015). Hierarchical Incident Ticket Classification with Minimal Supervision. Proceedings - IEEE International Conference on Data Mining, ICDM. 2015.
14. Pradhan, Dr & Bagbande, Arun & Khan, Amaan & Majid, Abdul & Chandekar, Unnati. (2022). ITSM Using AI Chat-Bot and Data Visualizers. International Journal for Research in Applied Science and Engineering Technology. 10. 704-708. 10.22214/ijraset.2022
15. https://processflows.co.uk/process-automation/
16. https://www.kaggle.com/datasets/nikolagreb/small-itsm-dataset

17. https://www.kaggle.com/
18. https://colab.research.google.com/
19. https://www.nvidia.com/en-in/data-center/tesla-t4