

Course Title: Research Methodology in Computer Science

Course Code: AI800

Weighting: 20%

Lecturer: Dr. Mohammed Anbar

Hand-out Date: 30 Oct 2025

Hand-in Date: 08 Nov 2025

Hand-in Method: Online Submission via E-Learning. Assignment Report:

A Survey of NLP Techniques for:

Automated IT Support Ticket Classification and Resolution

Student ID: 253001724

Student Name: Anas Abdulkader Alzain

Table of Contents

- Introduction
- Summary of Research Papers
- Critical Analysis
- Connections
- Connections Between the Studies
- Synthesis Matrix for Literature Review
- References

Introduction

Present day IT service management environments handle thousands of support tickets every day. Tickets range from simple access requests to system failures. Manual ticket classification and routing is error prone and time consuming leading to delays, mis routed issues and increases operational costs.

Researchers are applying NLP and ML to automate ticket management. This literature review examines six studies published between 2023 - 2025. It explores various methods for automated ticket classification, prioritization, routing and resolution. The objective is to identify current trends, evaluate effectiveness and highlight areas of improvement.

Summary of Research Papers

❖ Paper 1:

Ticket-BERT: Labeling Incident Management Tickets with Language Models (Liu et al., 2023)

This seminal work from Microsoft Research introduces Ticket-BERT, a specialized BERT-based model for automated ticket classification in large-scale enterprise environments. The paper addresses three critical challenges unique to ticket data: tickets are created by both automated systems and human engineers with varying protocols, descriptions undergo frequent revisions during the ticket lifecycle, and labeling requirements change rapidly as software and hardware evolve.

The authors developed three real-world datasets (D-Human, D-Machine, D-Mixture) containing ticket titles, summaries, and multiple description updates. Ticket-BERT extracts the [CLS] token embedding from BERT-base and feeds it to a multi-layer perceptron for classification. The model demonstrates superiority over Microsoft Cognitive Services and various baselines across all datasets.

A key innovation is the active learning cycle deployed on Microsoft's IcM system, which processes 7,000 new tickets daily. This enables continuous model refinement with minimal annotation burden. The system also uses auxiliary prompt-prefix text sampling to boost performance on human-created tickets by 12-15%.

However, the approach struggles with extremely imbalanced categories and requires substantial computation resources for fine tuning. Domain adaptation remains challenging when transferring to new IT environments.

<https://arxiv.org/abs/2307.00108>

❖ Paper 2:

Machine Learning Framework for End-to-End Implementation of Incident Duration Prediction (Ajit et al., 2023)

This paper presents a framework for predicting incident resolution time and addressing a critical ITSM performance indicator. Accurate time estimates enable better resource allocation, improve SLA compliance, and enhance customer satisfaction.

The framework encompasses data preprocessing (outlier removal, missing value imputation, distribution analysis), automated feature selection using statistical techniques, and comparative evaluation of multiple machine learning algorithms. The authors tested classical models (Random Forest, SVM, XGBoost) and neural approaches (LSTM, GRU) on real service desk data.

Results showed Random Forest achieved 94.32% accuracy for clearance time prediction, while XGBoost delivered a mean absolute error of 6.44 minutes for response time estimation. The study revealed that incident priority, affected service category, and historical resolution patterns were the strongest predictive features.

Limitations include dependency on high-quality historical data, difficulty handling unprecedented incident types, and reduced accuracy during periods of organizational change. The model also struggles with incidents requiring cross-team collaboration.

<https://arxiv.org/abs/2304.11507>

❖ Paper 3:

Performance Evaluation of Machine Learning Techniques on Resolution Time Prediction in Helpdesk Support System (Tai et al., 2024)

This comprehensive benchmarking study evaluates various machine learning approaches for helpdesk resolution time prediction using the UCI Incident Management Process dataset. The research provides practical guidance for organizations selecting prediction algorithms.

The method involves systematic data preprocessing, feature engineering with statistical selection methods, and rigorous model comparison across accuracy, training time, and inference speed metrics.

Models that were tested include Gradient Boosting, XGBoost, Support Vector Regression, K-Nearest Neighbors and neural networks.

The outcome is that Random Forest and XGBoost consistently outperformed single models.

However, the research is limited to English-language tickets and a single organizational context. The models show degraded performance for tickets with vague descriptions or when multiple issues are combined in one request.

<https://joiv.org/index.php/joiv/article/view/2305>

❖ **Paper 4:**

Enhancing IT Incident Management with Natural Language Processing and Predictive Analytics (Recent Study, 2025)

This cutting-edge research explores hybrid approaches combining NLP for text understanding with LSTM networks for temporal pattern recognition in incident management. The paper represents a paradigm shift from reactive to proactive IT operations.

The hybrid model uses NLP to parse and prioritize incoming tickets based on content analysis (keywords, sentiment, urgency indicators), while LSTM networks analyze historical incident sequences to predict potential failures before they occur. The integration of domain knowledge graphs provides richer context-aware understanding of incident relationships.

The outcome was 65% improvement in response efficiency when pitted against traditional rule-based systems. The model successfully predicted 78% of recurring incidents 24-48 hours before users reported them, enabling preventive maintenance.

Key drawbacks include the need for extensive labeled training data, challenges with rare incident types, and challenges handling cross-domain incidents. Privacy concerns also arise when analyzing detailed user behavior patterns. The federated learning approach proposed addresses some privacy issues but increases implementation complexity.

https://www.researchgate.net/publication/392465831_Enhancing_It_incident_management_with_natural_language_processing_and_predictive_analytics

❖ **Paper 5:**

A Machine Learning-Based Predictive Model for the Management of Incidents in Small and Medium-Sized Enterprises (Peru Study, 2024)

This ACM conference paper addresses the specific challenges of implementing AI-driven ITSM in resource-constrained SME environments. Unlike enterprise solutions requiring substantial infrastructure, this model focuses on practical deployment with limited computational resources.

The research proposes a supervised multi-label classification algorithm trained on historical incident data to automatically prioritize IT incidents. The model considers incident impact, urgency, affected systems, and user roles to assign priorities aligned with business criticality.

Testing on SME data from Peru demonstrated successful automation of incident prioritization with 85% accuracy using relatively small training datasets (5,000-10,000 tickets). This implementation is lightweight and runs efficiently on standard hardware without GPU acceleration.

The model's performance significantly degrades when ticket descriptions contain domain-specific jargon or are poorly written. It also struggles with incidents requiring subjective business judgment beyond technical categorization. The research highlights the need for region-specific training data to account for language variations and local business practices.

 <https://dl.acm.org/doi/10.1145/3654823.3654913>

❖ Paper 6:

Automated Customer Ticket Classification Using BERT (Inspirisys Case Study, 2023)

This industry case study demonstrates real-world BERT implementation for multi-class ticket classification in enterprise customer support. The paper provides valuable insights into production deployment challenges and practical solutions.

The solution fine tunes a pre-trained BERT model using preprocessed unstructured ticket text by removing html tags, special characters urls and ip addresses. This approach reduces training times and compute costs.

In production, the system can handle tickets in multiple languages It achieves 89% classification accuracy with under 200ms inference time per ticket, enabling real-time processing.

Limitations include difficulty with extremely short tickets (under 10 words), confusion when tickets describe multiple issues, and reduced accuracy for newly introduced product categories. The system requires periodic retraining as business offerings evolve. The system is integrated with existing ticketing systems through REST API implementations.

 <https://www.inspirisys.com/blog-details/Customer-Ticket-Classification-using-BERT/136>

Critical Analysis

All six studies tackle real-world IT support challenges but consistently highlight issues with data quality, model generalization, computational cost, and organizational adoption.

Papers 1 and 6 confirm that transformer-based models—especially BERT variants—now set the standard for ticket classification, outperforming traditional ML methods. Paper 1's Ticket-BERT showcases enterprise-scale capability with active learning, while Paper 6 proves that lightweight BERT fine-tuning is practical even for mid-sized organizations.

Papers 2, 3, and 5 shift focus toward predictive tasks like SLA management and resource planning, showing that ensemble models such as Random Forest and XGBoost deliver the best balance between accuracy and efficiency. Paper 3's benchmarking provides useful, implementation-ready insights.

Paper 4 aims to forecast failures before they occur. though promising, they still expose key hurdles which includes massive data demands and low accuracy for rare events.

Across all studies, four recurring style stand out:

High quality data is needed for good model performance; poor or inconsistent tickets lead to erroneous results.

Models often fail to generalize beyond the data they were trained on. black-box systems create trust barriers when handling critical issues.

Static models quickly fall behind in dynamic IT environments requiring continuous fine tuning.

Success depends on clean data, adaptable models, transparent reasoning, and continuous improvement.

Connections Between the Studies

The common factors that exist between these papers:

BERT and its variants (Papers 1, 6) have outperformed traditional NLP methods (TF-IDF, Word2Vec) for understanding ticket descriptions. The bidirectional context modeling of transformers is important for capturing domain-specific language in IT tickets.

Random Forest and gradient boosting (Papers 2, 3, 5) deliver the best results for predicting resolution time, priority, and routing decisions based on structured ticket metadata.

Papers 1 and 4 combine text analysis, metadata features, and domain knowledge to achieve superior performance. Single-modality approaches leave valuable information unused.

Given rapid IT environment changes (Paper 1), static models quickly become stale. Continuous model updates with human-in-the-loop validation represent best practice for production systems.

Unlike purely academic exercises, these papers emphasize real-world constraints: inference speed (Paper 6), limited training data (Paper 5), integration with existing systems, and computational efficiency.

Conflicts and Contradictions:

Although most studies agree on the main goals, they differ in several important areas:

- Paper 1 uses an encoder-only BERT model, while Paper 4 argues that encoder-decoder or decoder-only models may handle errors better. There's still no clear answer on which type works best for different IT support tasks.
- Papers 2 and 3 focus on manually selecting useful features, while Papers 1 and 6 depend on BERT to learn everything automatically. The right balance between the two is still uncertain.
- Paper 2 measures success mainly by classification accuracy, while Paper 3 uses MAE (Mean Absolute Error) for prediction tasks. Few papers evaluate real business impacts like SLA compliance or cost reduction.
- Paper 5 reports good results using around 5,000 training examples, while Paper 1 used much larger datasets. It's still unclear how much data is truly needed for reliable model performance.
- Paper 6 accepts BERT's "black box" nature to gain higher accuracy, but Papers 3 and 5 prefer models like Random Forest that are easier to explain. More work is needed to decide how important explainability really is for IT teams.
- Only Paper 4 discusses privacy through federated learning. The rest overlook this issue, even though IT tickets often contain sensitive organizational data.

Synthesis Matrix for Literature Review

Paper	Year	Author(s)	Methodology	Key Findings	Weaknesses
Ticket-BERT	2023	Liu et al. (Microsoft)	BERT-base fine-tuning; active learning; 3 datasets (D-Human, D-Machine, D-Mixture); auxiliary prompt-prefix	Outperforms Microsoft Cognitive Services; handles 7K daily tickets; 12-15% improvement with auxiliary text	High computational cost; domain adaptation challenges; requires substantial labeled data
Incident Duration Prediction	2023	Ajit et al.	End-to-end ML framework; Random Forest, SVM, XGBoost, LSTM, GRU; UCI dataset	Depends on quality historical data; struggles with novel incidents; poor cross-team collaboration prediction	framework; Random Forest, SVM, XGBoost, LSTM, GRU; UCI
Resolution Time Evaluation	2024	Tai et al.	Comprehensive benchmarking; Linear Regression, Decision Trees, RF, XGBoost, SVR, KNN, Neural Networks	XGBoost optimal balance (6.44 min MAE + efficiency); ensemble methods best; identifies critical features	English-only; single organization; degrades with vague descriptions; multi-issue tickets problematic
NLP + Predictive Analytics	2025	Recent Study	Hybrid NLP-LSTM; domain knowledge graphs; temporal pattern recognition; federated learning	65% efficiency improvement; 78% recurring incident prediction (24-48h advance); proactive maintenance enabled	Extensive labeled data required; rare incident challenges; cross-domain struggles; privacy concerns; complex implementation
SME Incident Prioritization	2024	Peru Study (ACM)	Supervised multi-label classification; lightweight deployment; 5K-10K training samples	85% accuracy; SME-appropriate; runs on standard hardware; no GPU required	Degrades with jargon/poor writing; struggles with subjective judgment; needs region-specific data
BERT Classification (Industry)	2023	Inspirisys	Pre-trained BERT fine-tuning (3-5 epochs); multi-language support; production deployment	89% accuracy; <200ms inference; real-world integration; minimal fine-tuning	Struggles with short tickets (<10 words); multi-issue confusion; new category challenges; requires periodic retraining

References

- Liu, Z., Wingate, D., & Shou, Z. (2023). Ticket-BERT: Labeling Incident Management Tickets with Language Models. *arXiv preprint arXiv:2307.00108*.
<https://arxiv.org/abs/2307.00108>
- Ajit, S., Srivastava, P., Sharma, R., & Kumar, A. (2023). Machine learning framework for end-to-end implementation of Incident duration prediction. *arXiv preprint arXiv:2304.11507*.
<https://arxiv.org/abs/2304.11507>
- Tai, T.-E., Haw, S.-C., Kong, W.-E., & Ng, K.-W. (2024). Performance Evaluation of Machine Learning Techniques on Resolution Time Prediction in Helpdesk Support System. *International Journal on Robotics, Automation and Sciences*, 6(2), 59-68.
<https://journals.mmupress.com/index.php/ijoras/article/view/1078>
- Tai, T., Haw, S., Ng, K., Al-Tarawneh, M., & Tong, G. (2024). Performance Evaluation on Resolution Time Prediction Using Machine Learning Techniques. *JOIV: International Journal on Informatics Visualization*, 8(2), 583-591.
<https://joiv.org/index.php/joiv/article/view/2305>
- Recent Authors (2025). Enhancing IT incident management with natural language processing and predictive analytics. *International Journal of Science and Research Archive*, 15(3), 224-237.
https://www.researchgate.net/publication/392465831_Enhancing_It_incident_management_with_natural_language_processing_and_predictive_analytics
- Peru Study Authors (2024). A Machine Learning-Based Predictive Model for the Management of Incidents in Small and Medium-Sized Enterprises in Peru. *Proceedings of the 2024 3rd Asia Conference on Algorithms, Computing and Machine Learning (CACML '24)*. <https://dl.acm.org/doi/10.1145/3654823.3654913>
- Inspirisys (2023). Automated Customer Ticket Classification using BERT. *Inspirisys Blog*.
<https://www.inspirisys.com/blog-details/Customer-Ticket-Classification-using-BERT/136>