JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AUTOMATIC IT TICKET ASSIGNMENT

[1]**Prof. Anand S Pashupathimath**,[2]**Ankita V Gulaganji**, [3]**Apeksha K Shiraguppi**, [4]**K Bhoomika**, [5]**Samiya ShaikSandi**
[1]**Professor,** [2]**Student ,** [3]**Student ,** [4]**Student,** [5]**Student**

Department of Computer Science and Engineering
SDM College of Engineering and Technology, Dharwad, India

Abstract: In modern IT industries, Incident Management plays a major role in ensuring efficient customer support. Traditionally in IT industries, assigning incoming tickets to relevant teams remains a manual process, leading to human errors, resource inefficiencies, and longer response and resolution times, ultimately lead to poor customer service. This project aims to address these challenges by developing a Al-based classifier model for text classification that automatically assigns incoming tickets to the appropriate Assignment groups based on their descriptions and also aiming to improve cost effectiveness and reduce resolution times. Our approach involves comprehensive data preprocessing techniques, including cleaning, translation, tokenization, and lemmatization. We assess the performance of different machine learning algorithms, such as Bi-gram models, Gensim LDA, Word2Vec, GloVe embeddings, Bidirectional LSTM, Random Forest, and SVM classifiers.In summary, this research offers a comprehensive approach to incident management, leveraging NLP and ML techniques to automate incident analysis and allocation.The proposed solution demonstrates our evaluation, based on training and test accuracies, shows that the LSTM model trained on resampled data achieves the highest accuracy of 87%. Our research contributes to enhancing incident management practices in the IT industry.

***Key words**: Al based classifier model, data preprocessing RF,SVM,LSTM.*

1. INTRODUCTION

Organizations today mostly depend on IT service for resolving technical issues and maintaining smooth operations of IT System and services,which highlights the vital role that IT service desks play in the corporate world.

The main duties of a service desk team typically include handling incoming incident tickets, providing technical support to users, troubleshooting issues, and escalating complex problems to specialized teams when necessary. In the past, organizations have assigned tickets manually, and this has increased resource usage and lengthened query response times, which has eventually lowered customer satisfaction.

So, the objective of our project is to automate the incoming incident assignment to the right functional group by analyzing the given description. This automation aims to reduce costs and time significantly while improving response time and efficiency. To achieve this goal, we built models that leverage machine learning techniques. Through result analysis, we concluded that the LSTM model provides higher accuracy in ticket assignment.

2. LITERATURE REVIEW

[1] "**TaDaa : Ticket Assignment Deep learning Auto Advisor**" by Yang, L., Liu, Z., & Li, X. (2022):This paper presents TaDaa, a deep learning-based auto ticket assignment system using a dataset of customer support tickets. The system uses a hierarchical classification approach, with the first level classifying tickets into broad categories and the second level assigning tickets to specific support teams.

[2] "**Automatic Ticket Assignment using NLP**" by Siva, R. (2020): In this paper, we used machine learning models like Stochastic Gradient Descent (SGD), Multinomial Naive Bayes (NB), Linear Support Vector Machine (SVM), and Logistic Regression to classify L1/L2 tickets into GRP0 and GRP8. Another model was developed for L3 ticket classification.

[3] "**Automatic-IT-Ticket-Assignment-NLP":**This paper aims to automate IT ticket assignment using an AI-based classifier model, targeting an accuracy threshold of 85%. Data preprocessing included text cleaning, translation, stop word removal, and lemmatization,

with word embeddings generated using word2vec and GloVe. Various NLP algorithms like LSTM RandomForest, and SVM were applied for classification. One limitation involves grouping low-entry assignment groups,

potentially requiring manual intervention.

3. SYSTEM OVERVIEW :

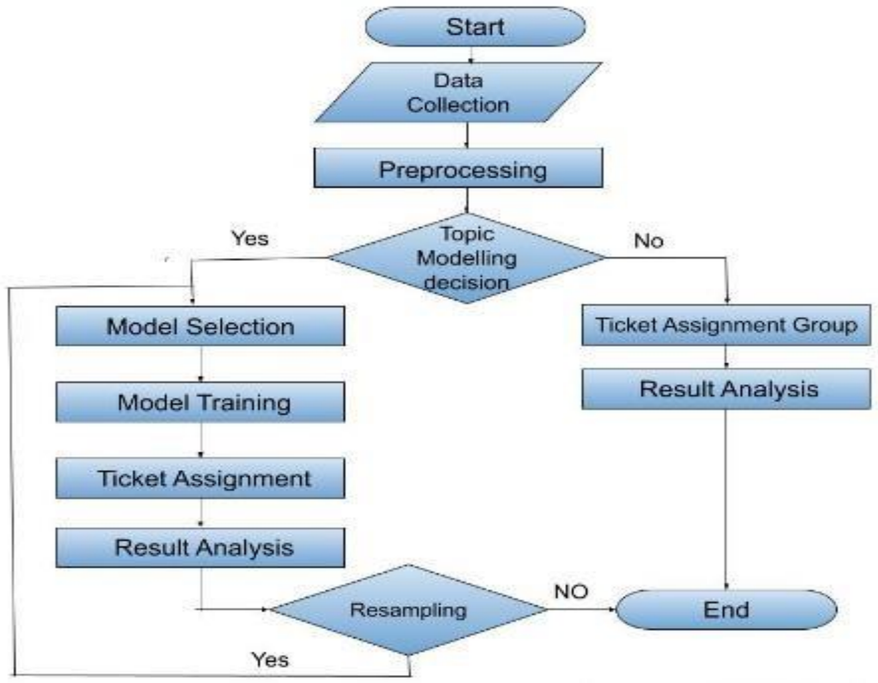

Fig 1: Overall Ticket Assignment Process

The flowchart depicts the automated ticket assignment process. First, data is collected, likely from sources like Kaggle. This data undergoes pre-processing to ensure quality, including cleaning, translation , and lemmatization. A crucial step follows: the "Topic Modeling Decision." Here, the system analyzes the ticket description using Natural Language Processing to see if it aligns with a predefined topic (e.g., network issues). If there's a clear match and the pre-trained model is confident handling this topic, the ticket goes for automated assignment. Then the process undergoes Model Selection and it is trained on relevant data before assigning the ticket to the most suitable group. The model's performance is then analyzed, with the option to resample the data if needed for improved accuracy and to balance the workload. If the ticket description doesn't align with a predefined topic or automated assignment is deemed unsuitable, the ticket is manually assigned by a team lead or manager, considering factors like agent expertise and urgency. This process ensures efficient ticket assignment through automation for clear-cut cases, while also allowing for human intervention for complex or ambiguous situations.

4. RESEARCH METHODOLOGY:

1. Data Analysis: The dataset contains four columns: Short Description, Description, Caller, and Assignment Group, total of 74 distinct Assignment Groups categorized as target classes. Caller names are arranged in a random order, lacking relevance for training data analysis. Furthermore, the dataset includes instances of European non-English languages, in addition to symbols and various characters embedded within the descriptions. Some descriptions closely resemble their corresponding short descriptions. Additionally, spelling mistakes and typo errors are also found.

2. Data Pre-processing: The preprocessing steps included dropping irrelevant fields, handling missing values, merging text fields, and standardizing case sensitivity. Language translation was applied for consistency, while tokenization and stop word removal streamlined the text. Lemmatization and spell checks ensured data accuracy, and WordClouds visualized key terms for each Assignment Group, complemented by distribution plots for data pattern insights.

3. Data Cleaning: The clean_data() function standardizes the text by converting all words to lowercase, ensuring consistency in analysis. This process effectively addresses variations arising from different word cases, streamlining further data processing and analysis task.

4. Translation: Non-English text, particularly in German, was identified in the dataset.Various libraries such as googletrans, textblob, goslate, etc., were attempted for translation but encountered size limitations. A wordlist of non-English words from the dataset was created to overcome translation limitations. Filtering the Description column using the German wordlist, non-English entries were translated to English using Google Translate.Translation aimed to improve model accuracy by ensuring a consistent language format.

5. Lemmatization and Stop Words Removal: In Tokenization process the text data was tokenized into individual words to prepare

for lemmatization. In Lemmatization each word was lemmatized using NLTK's WordNetLemmatizer, which maps words to their base forms based on their part of speech. Stop Words Removal: Additionally, common stop words such as 'is', 'the', 'and', etc., were removed to focus on meaningful content

6. Word Clouds: Word Cloud is a data visualization technique which represent the frequency of word occurrences in a dataset . By analyzing word clouds for the top 50 words in each group, we observed distinct patterns and associations.

7. Handling Dataset Imbalance: Imbalance in a dataset can lead to biased models that perform well on the majority class but poorly on minority classes. Through analysis we found that Group0 needs to be downsampled, while all other groups should be upsampled.After performing resampling on the dataset this step play a siginifcant role in model evaluation.

8. Word Embedding: Word2Vec and GloVe are popular techniques used for word embedding, converting textual data into numerical representations. This step is crucial for training deep learning models.

9. Model Training: The dataset is spilt two models 80% for training and 20% for testing using train_test_split libraray. In training, the model learns from labeled data to optimize its parameters for making predictions.During testing, the trained model's performance is evaluated on unseen data to assess its predictive accuracy.Validation involves tuning the model's hyperparameters and assessing its performance on a separate dataset to prevent overfitting.

10. Model Selection: We explored different models like Bi-directional LSTM, Random Forest, and SVM for ticket assignment and functional group categorization. Each model was designed with care, considering factors like layer depth, activation functions, and dropout regularization to prevent overfitting and ensure effective performance.

11. Model Evaluation: The models are evaluated using a variety of metrics such as Precision, Recall, F1-score, Confidence Interval, and Accuracy.

12. Result Analysis: Result is analyzed based on the performance of each model and the result is been captured.

5. Experimental results and Discussion:

1. Performance Evaluation: The performance of the Bi-directional LSTM models was notable, achieving accuracies of 87% on the resampled augmented dataset. These results demonstrate the effectiveness of deep learning models in text classification tasks, surpassing the performance of traditional ML algorithms like Random Forest and SVM.

2. Comparison to Benchmark: Comparing our AI-based classifier to the existing system's 75% accuracy in ticket assignment, the LSTM model excelled with an accuracy of 87%, meeting and exceeding our objective of reaching at least 85% accuracy. This improvement emphasizes the potential of advanced NLP techniques in enhancing IT service desk operations.

3. Discussion :The experimental results validate that employing deep learning models leads to a more effective approach in assigning incidents to the right functional group, resulting in lower costs, reduced time, and higher efficiency. This approach represents a significant advancement compared to manual processes, especially when dealing with real-world datasets and leveraging various machine learning models, ultimately leading to improved results and user satisfaction.

6. Result:

6.1 Result of Resampling the Dataset

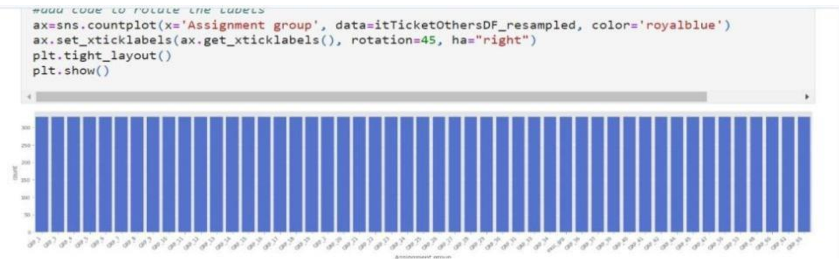

Fig.2 : Class Balancing

Here the above result explains the need of class balancing ie resampling the dataset The code appears to be using a combination of upsampling and downsampling.First,it upsamples the minority class by duplicating examples from the 'itTicketGrpDF' DataFrame. Then,it downsamples the majority class by removing examples from the 'itTicketOthersDF' DataFrame.

Following data resampling, the code generates a countplot illustrating class distribution across the dataset's assignment groups. This plot indicates ticket counts for each group. Resampling offers a crucial advantage by enhancing model performance, effectively minimizing inaccuracies in accuracy evaluations caused by misleading imbalances in class distribution.

6.2 Result of Accuracy of the Model

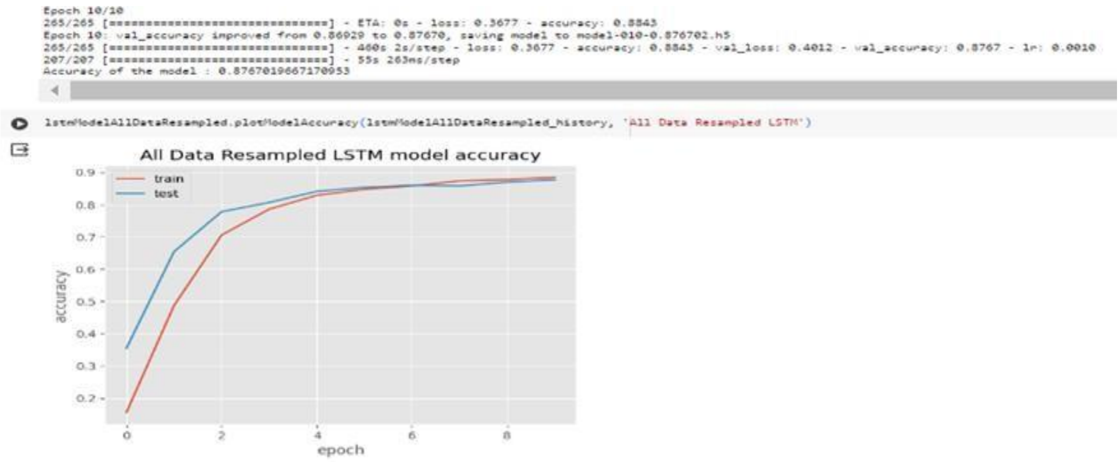

Fig.3 Accuracy of the model

Understanding the Plot:

1) The x-axis typically represents epochs, which are full cycles of the training data being fed through the neural network.
2) The y-axis on the left represents accuracy (usually between 0 and 1, where 1 indicates perfect accuracy).
3) The y-axis on the right represents loss (typically non-negative, with lower values indicating better model performance).

As the model trains, its accuracy (red line) ideally increases, indicating it's learning to correctly classify examples. Simultaneously, the loss (blue line) should decrease, signifying the model is making better predictions.

7. Conclusion :

The development and evaluation of the automated ticket assignment system using advanced NLP and ML techniques signify a substantial progression in IT industry. This research demonstrates the feasibility and efficiency of leveraging AI-based classifiers for streamlining incident handling processes.

Our comprehensive evaluation of the system's performance reveals its capability to accurately categorize a diverse range of IT tickets, including issues related to password resets, network outages, system failures, and software glitches. By employing various machine learning algorithms such as Bi-directional LSTM and GRU models, we have achieved high accuracy in assigning tickets to the appropriate functional groups.

Our system offers several advantages over traditional manual assignment methods, including enhanced accuracy, reduced resolution times, and optimized resource utilization. So in comparison with manually our system approach proves to be better and can solve all the above mentioned issues .

However, it is important to acknowledge certain limitations of our research, such as the need for continuous monitoring and model retraining to adapt to evolving IT landscapes. Challenges related to data sparsity, multilingual inputs, and outlier detection also require ongoing attention for maintaining optimal system performance

Future Scope

In the future, researchers may explore deep reinforcement learning for dynamic ticket assignment strategies. They may also incorporate sentiment analysis to improve customer experience, and integrate chatbot interfaces for interactive incident management. The study contributes to advancing incident management practices in the IT industry by leveraging cutting-edge NLP and ML technologies. The successful implementation of automated ticket assignment systems can lead to improved service quality, reduced operational costs, and enhanced customer experiences.

In summary, this study contributes to advancing incident management practices in the IT industry by leveraging cutting-edge NLP and ML technologies. The successful implementation of automated ticket assignment systems can lead to improved service quality, reduced operational costs, and enhanced customer experiences, positioning organizations for greater efficiency and competitiveness in the digital era.

8. References:

[1] "Automatic Ticket Assignment using NLP" by Siva, R. (2020): https://www.kaggle.com/datasets/aviskumar/automatic-ticket-assignment-using-nlp

[2] "TaDaa: Ticket Assignment Deep learning Auto Advisor" by Yang, L., Liu, Z., & Li, X. (2022): **https://arxiv.org/abs/2207.11187**

[3] "Auto Ticket Assignment Application using NLP " by Divya R Kamat (2020): **https://github.com/divya-r-kamat/AutoTicketAssignment**

[4] "AUTO TICKET ASSIGNMENT TO THE RIGHT FUNCTIONAL GROUPS USING MACHINE LEARNING AND RPA TECHNIQUES | by Great Learning Snippets | Medium" **https://medium.com/@blogsupport/auto-ticket-assignment-to-the-right-functional-groups-using-machine- learning-and-rpa-techniques-fcd59ea87d4c**