In23-S1-CS5617 - Data Science

Analysis & Predicting MTTR on Incident Response Log

239340T – Nandasena R.B.M.S.

University of Moratuwa

2023

Table of Contents

List of Tables

List of Figures

1. INTROUDCUTION

Incident response log analysis is important for several reasons:

- **Identifying the root cause of incidents**: Incident response logs can provide valuable information about the events that led up to an incident, making it easier to identify the root cause of the problem.
- **Improving incident response processes**: Analyzing incident response logs can help organizations identify areas where their incident response processes can be improved. This can include everything from how incidents are detected and reported, to how they are escalated and resolved.
- **Learning from past incidents:** Analyzing incident response logs can provide valuable insights into past incidents, including what worked well and what did not. This information can be used to develop best practices and improve incident response processes for the future.

1.1 Detail description of the selected dataset

The dataset is a response log of an incident management process obtained from the audit system of a ServiceNow platform utilized by an IT company. It contains 36 attributes, including:

number, incident_state, active, reassignment_count, reopen_count, sys_mod_count, made_sla, caller_id, opened_by, opened_at, sys_created_by, sys_created_at, sys_updated_by, sys_updated_at, contact_type , location, category, subcategory, u_symptom, cmdb_ci, impact, urgency, priority, assignment_group, assigned_to, knowledge, u_priority_confirmation, notify, problem_id, rfc (request for change), vendor, caused_by, close_code, resolved_by, resolved_at, closed_at

Data Source: https://www.kaggle.com/datasets/vipulshinde/incident-response-log?datasetId=771310&sortBy=dateRun&tab=collaboration

1.2 Research Questions

- How does the MTTR change over time for different types of incidents?

  Here we must build a model capable of predicting the MTTR based on the several key attributes which we can think directly affect and decide the MTTR. Below listed attributes are the key values deciding the MTTR.

  - Impact wise
  - Priority wise
  - Urgency wise
  - Incident State wise

- Can we use machine learning models to accurately predict the MTTR for new incidents based on the attributes in the dataset?

  Since we are aware of the x, and y values this needs to be addressed using supervised learning. Therefore, I have chosen below ML algorithms to build and evaluate the MTTR. The final outcome of this is to find the best-fitted model to predict the MTTR.

  - Linear regression
  - KNN
  - Random forest
  - Decision tree

1.4 Expected Deliverables/ Outcomes

MTTR stands for Mean Time to Repair, which is the average time it takes to repair a incident after a failure it has been raised. In incident management, MTTR is a key metric used to measure the effectiveness and efficiency of the incident response process. The lower the MTTR, the better the incident response process is considered to be.

Addressing the mentioned research questions, would be provided valuable insights into the incident resolution process and identify areas for improvement, optimize resource allocation, incident prioritization, capacity planning, and provide better service to customers.

2. DATA PREPARATION

2.1 Data Transformation

In data transformation, I created 1 attribute which is time_to_resolve and group according to the above-mentioned key attributes.

Here we are converting the 'opened_at' and 'resolved_at' columns to datetime format, calculated the MTTR (Mean Time to Resolve) in minutes, and removed rows with negative or zero MTTR values. These transformations help to convert the raw data into a format suitable for analysis and make it easier to extract meaningful insights from the data.

Time_to_resolve is the time it takes to repair once incident has been raised. After grouping by the selected attributes, we need to take the mean values for each group and that should be kept in the variable called MTTR which means Mean time to resolve.

3. DESCRIPTIVE ANALYSIS

In this step what we are doing is, plot several boxplots that showing the relationship between the incident state and the MTTR. The boxplot for each incident state is plotted with MTTR on the y-axis. This helps to identify if there are any outliers or if there is a significant difference in MTTR across different incident states. Similarly, hue parameters such as 'impact', 'urgency' and 'priority' are used to group and visualize the relationship between those parameters and MTTR.

After exploring the relationships between the variables through the boxplots, the next step is to visualize the correlation between MTTR and variables. We are doing this using a correlation matrix and heatmap. The correlation matrix shows the correlation between each pair of variables, including the correlation between MTTR and variables. This helps to identify which variables are strongly correlated with MTTR and which are weakly correlated.
As per the observations below is the heatmap.

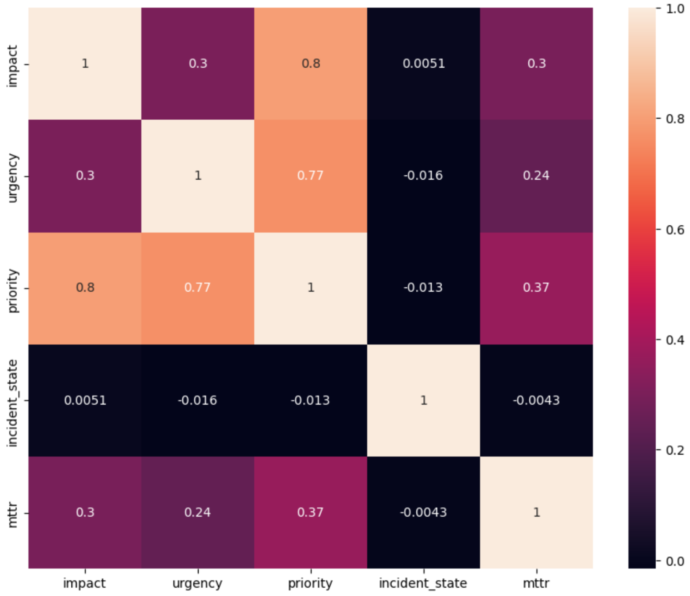
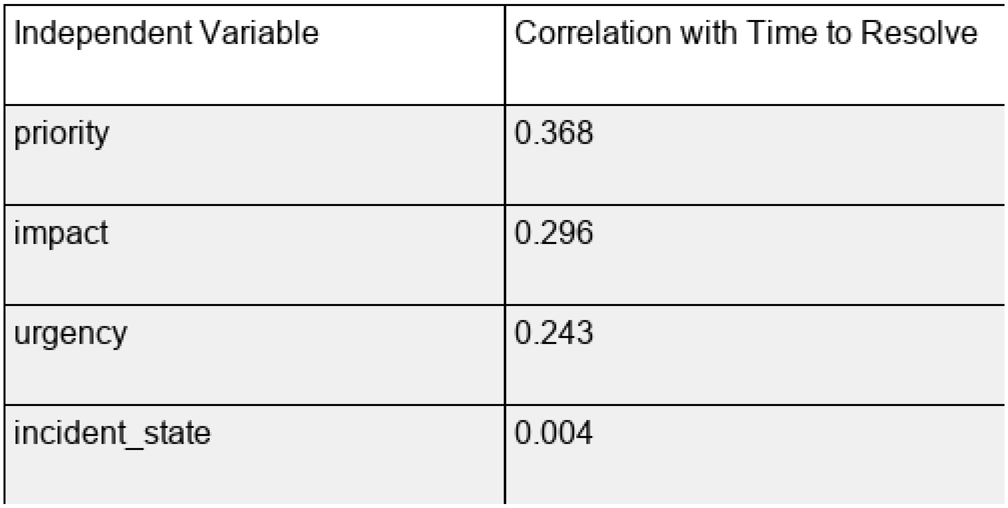


Figure 1: Correlation between all the categorical attributes in the dataset



Independent Variable	Correlation with Time to Resolve
priority	0.368
impact	0.296
urgency	0.243
incident_state	0.004

Table 1: Correlations with the time to resolve vs. key attributes

From the table, we can see that the priority of an incident has the highest correlation with the time to resolve (MTTR) at 0.369, followed by impact at 0.296 and urgency at 0.243. The incident state, on the other hand, has a very low correlation with MTTR at 0.004.

This suggests that the priority, impact, and urgency of an incident are important factors in determining how long it takes to resolve an incident. In other words, incidents with higher priority, impact, and urgency tend to take longer to resolve. However, incident state does not seem to have a significant impact on MTTR, at least in the current dataset.

3. PREDICTIVE ANALYSIS

In the predictive analysis, we tried to build a model to predict the MTTR (Mean Time to Resolve) based on the independent variables impact, urgency, and priority. We used four different machine learning algorithms to build our models: Linear regression, KNN, Random Forest, and Decision tree.

To evaluate the performance of each model, we used R-squared as our metric. R-squared is a statistical measure that represents the proportion of variance in the dependent variable (MTTR in our case) that can be explained by the independent variables (impact, urgency, and priority). The value of R-squared ranges from 0 to 1, where 0 means the model does not explain any variance in the dependent variable, and 1 means the model explains all the variance in the dependent variable.

The R-squared values for each model were as follows:

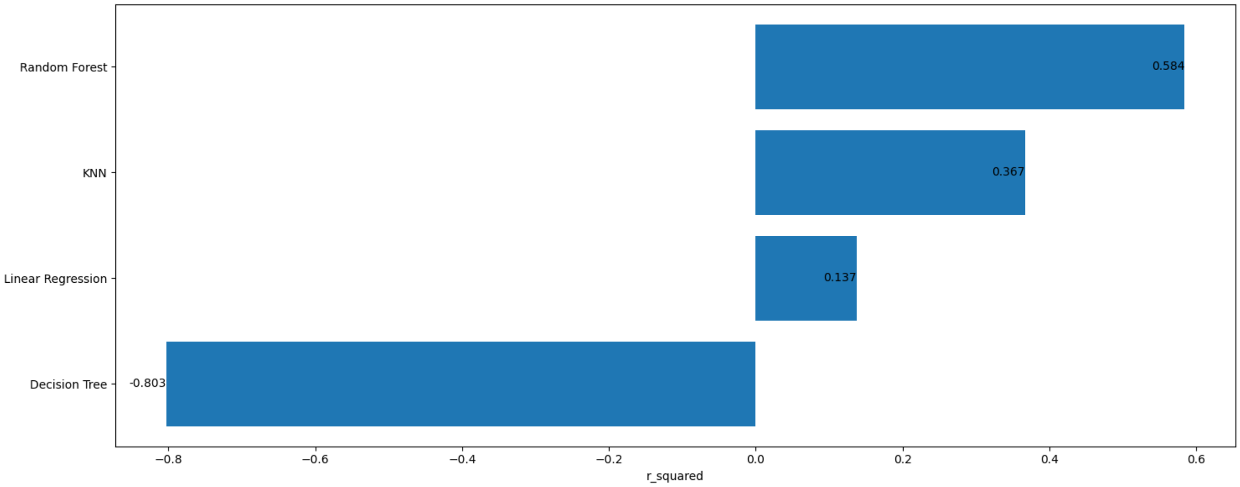
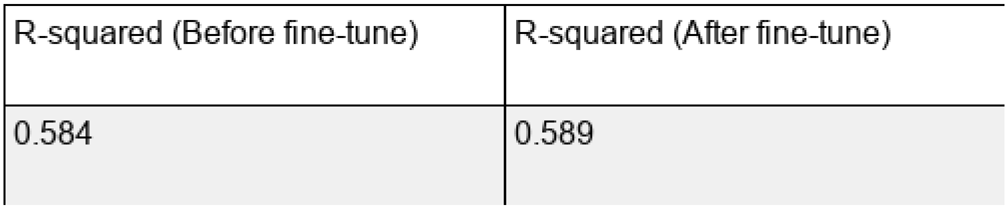


Figure 2: R-squared values for each model

- Decision Tree: -0.803
- Linear Regression: 0.137
- KNN: 0.367
- Random Forest: 0.584

Based on these results, we found that the Random Forest model performed the best among the four models, with an R-squared value of 0.584. This means that the independent variables (impact, urgency, and priority) in our model explain 58.4% of the variance in the MTTR.

To further optimize the performance of the Random Forest model, we are tuning the hyperparameters of the model using GridSearchCV. By doing so, we can fine-tune the model and improve its accuracy in predicting the MTTR.



R-squared (Before fine-tune)	R-squared (After fine-tune)
0.584	0.589

Table 2: Fine-tune hyperparameters for random forest using grid search and cross-validation

4 CONCLUSION & RECOMMENDATIONS

4.1 Conclusion

Based on the analysis and predictions made in this scenario, there are several valuable insights that decision-makers can gain.

Firstly, they can identify the factors that have the greatest impact on the time it takes to resolve an incident. The analysis shows that priority, impact, and urgency are the most significant factors affecting MTTR. Therefore, decision-makers could focus on improving the process of resolving high-priority incidents, which would have a significant impact on reducing MTTR.

Secondly, they can use the predictive models developed to estimate the expected MTTR for a new incident based on its priority, impact, and urgency. This information can be used to set realistic expectations for stakeholders and manage their expectations accordingly.

Thirdly, decision-makers can use the analysis to identify areas of improvement in the incident management process. For example, they could analyze the factors contributing to incidents with longer MTTRs and identify areas for improvement. They could also analyze the correlation between MTTR and other variables, such as reassignment count and sys_mod_count, to identify any patterns that could indicate areas for improvement.

4.2 Recommendations

To improve the incident response process and reduce MTTR, the following will be recommend based on the predictions:

1. Conduct regular training sessions for incident response teams to improve their skills and knowledge.
2. Develop a comprehensive incident response plan that includes clear guidelines for incident categorization and prioritization.
3. Implement a system for real-time incident monitoring to identify and resolve incidents more quickly.
4. Conduct regular reviews of incident response processes to identify areas for improvement and implement necessary changes.