ITSM Predictive Analytics Dashboard

From Reactive Incident Management to Predictive Service Excellence

Prepared by: Kartik Kodilkar | Business Analyst | ITIL v4 Certified | MSc Business Analytics (First Class Honours, Dublin Business School 2025)
Contact: Kartik Kodilkar | +353 83 854 7122 | kartikkodilkar.org | July 2026

Executive summary

Overview

IT service management teams managing incident operations at scale, there is the challenge of systematic prioritization as opposed to a problem of resource availability. This project demonstrates how the use of machine learning on ITSM professional data can move incident management from reactive state into proactive service excellence.

The project scope includes end to end analytical delivery, including data generation in python based on real patterns of enterprise incidents, feature engineering, modelling and evaluation of machine learning models in three algorithms. In addition to that, there is a three pages interactive powerBI dashboard delivering breach prediction insights to both operational teams and senior management. There were three models considered for this project, Random Forest, Logistic Regression and Gradient Boosting. The Random Forest model was selected with the accuracy of 71.97% and 80.61% of ROC AUC. Identifying 391 high-risk incidents and surfacing EUR 4.5M as annual exposures for SLA breaches in 24,918 incidents. A 36.58% SLA breach rate should not be viewed as a problem of resource but as a problem of information which is sloved through predictive analysis.

Key Metrics	Value	Implications
Total Incidents Analysed	24,918	Statistically real dataset for ML model training and evaluation.
Total SLA Breaches	9,115	36.58% SLA breach rate operationally unacceptable in most SLA contracts.
SLA Compliance Rate	63.42%	Below industry benchmark of 90 to 95% compliance

High Risk Incidents Identified	391	Actionable list for immediate proactive intervention
Avg. Breach Probability (High Risk)	45.12%	Sufficient confidence for operational prioritisation decisions
Random Forest ROC AUC	80.61%	Strong discriminative ability model reliably separates breach from non breach

Key Findings

1. Key Findings Volume-Based Prioritisation is Systematically Wrong

The category 46 ranks fourth in terms of number of incidents, but has the highest share of SLA violations, which means the violation rate is 51%. An approach for resource allocation based on the volume of incidents will put more emphasis on category 42, which has the largest volume but the lowest failure rate. From the data, it is evident that such a resource allocation strategy is ineffective. This trend would go unnoticed without the use of predictive analytics.

Category	Total Incidents (Rank)	Total SLA Breaches (Rank)	Implied Breach Rate
Category 42	3,600 (1st)	722 (4th)	~20%
Category 26	3,300 (2nd)	945 (3rd)	~29%
Category 53	2,700 (3rd)	985 (2nd)	~36%
Category 46	2,200 (4th)	1,114 (1st HIGHEST)	~51%

It is category 46 which contains the key lesson: companies allocate more resources to the wrong category, as there is no correlation between the size and the risk. Just that one lesson is worth all the analysis performed: moving the resource from category 42 to category 46 will result in decreased breach rate in several weeks

2. Key Findings Time of Day is an Unmanaged Breach Risk Factor

Open hour is one of the top four important predictive attributes with a share of 5.66%. The timing of the logged incident turns out to be a significant SLA violation predictor regardless of the nature and priority level of the incident. Logging of incidents outside working hours entails additional SLA breach risk which cannot be predicted by the current staffing policy. The risky times are consistently under-staffed since this risk remains undetected in current reports.

Operational Implication: Currently, organizations have a staffing policy which is the same for all times. However, according to the data, some times generate more SLA breach risk. It should not be solved with increased staffing but with time-sensitive resourcing.

3. Key Findings Agent-Level Variance is Driving Undetected Breach Risk

The opened_by_freq variable is the next most important predictor, contributing 10.83% to the model's feature importance. Indeed, the party responsible for documenting the case is one of the good predictors of an SLA breach, which could be indicative either of a training problem that has gone unnoticed or of a workload problem that does not get reflected in current KPI reports.

Operational implication: Some specific individuals or groups of people have higher rates of breaches than others. This phenomenon results from a training or workload problem and not from any performance problem per se. The model allows us to measure this variability for the first time.

Solutions

Solution 1 Random Forest Breach Prediction Model

Three classification models were trained and evaluated to predict SLA breach at individual incident level:

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Random Forest	71.97%	59.36%	74.11%	65.92%	80.61%
Logistic Regression	68.20%	54.20%	70.30%	61.20%	75.40%
Gradient Boosting	69.77%	56.80%	71.20%	63.10%	77.80%

The selection of Random Forest is due to the better results achieved for all measured criteria. The Recall rate of 74.11% is the crucial criterion which means that the system is able to correctly detect around 74% of breaches. In terms of ITSM, the cost of false negative decisions is higher than the cost of false positives. For each incident, the model produces the probability of breach along with the risk level (High, Medium, Low).

Solution 2 Risk-Based Incident Prioritisation Output

The Machine Learning Dashboard surfaces a ranked daily list of highest-risk open incidents by breach probability:

Incident ID	Breach Probability	Risk Category	Predicted SLA Breach	Actual SLA Breach
INC0000378	95.53%	High Risk	Yes	Yes
INC0000375	95.53%	High Risk	Yes	Yes
INC0000324	92.43%	High Risk	Yes	Yes
INC0000388	87.81%	High Risk	Yes	Yes
INC0000330	87.52%	High Risk	Yes	Yes
INC0000260	86.34%	High Risk	Yes	No
INC0000398	86.62%	High Risk	Yes	No
INC0000094	83.59%	High Risk	Yes	Yes
INC0000278	83.20%	High Risk	Yes	No

Predicted high risk events that did not cause any breach include INC0000260, INC0000278, and INC0000398. Being overly aggressive about an event that turns out to be harmless is much less costly than failing to detect a breach. Therefore, it would be more efficient for the model to err on the side of caution.

Solution 3 Three-Page Stakeholder-Aligned Power BI Dashboard

Dashboard Page	Target Audience	Key Decisions Supported
Executive Overview	Senior Management, Service Delivery Directors	Breach rate (36.58%), compliance rate (63.42%), top categories by volume and SLA breach strategic investment and governance decisions
Operational Analytics	Analytics, Incident Managers, Service Desk Team Leads	Category-level breach analysis resource reallocation from high-volume to high-risk categories including Category 46

Machine Learning Dashboard	Process Improvement, IT Analytics, Management	Individual risk scores, model comparison, feature importance proactive escalation and root cause identification

Recommended Impact

Recommendation	Solution Effectiveness	Impact	Notes
Rec 1 Risk-based prioritisation using RF model	Highly Effective	Redirects intervention to 391 high-risk incidents before breach estimated 10 to 20% breach reduction	RF ROC AUC of 80.61% provides reliable discriminative power for operational use
Rec 2 Category 46 resource reallocation	Highly Effective	Directly addresses the highest-risk category 51% breach rate vs 20% average	Immediate impact no technology investment required, resource reallocation only
Rec 3 Time-of-day and agent-level monitoring	Marginally Effective in isolation Highly Effective combined with Rec 1 and 2	Addresses 16.49% of total predictive feature importance (open_hour + opened_by_freq)	Requires workforce management integration and HR engagement for agent-level intervention

ANALYSIS

1. Research Methods

Method	Tool	Purpose
Data Generation	Python (Pandas, NumPy)	Synthetic ITSM dataset of 24,918 incident records modelled on real enterprise patterns
Feature Engineering	Python (Scikit-Learn)	Categorical encoding, frequency encoding and scaling for ML ready

		features
Model Training	Scikit-Learn (RF, LR, GB)	Supervised binary classification breach vs non-breach prediction
Model Evaluation	ROC AUC, Precision, Recall, F1	Performance validation Recall prioritised given cost asymmetry between false negatives and positives
Visualisation	Microsoft Power BI + DAX	Three page stakeholder aligned dashboard with dynamic KPI measures
Version Control	GitHub	Code repository, documentation and business case publication

2. Approaches Used

ITIL v4 Service Management Framework

The project is based on the ITIL v4 framework Continual Service Improvement, Problem Management and Service Level Management. The methodology follows the ITIL approach of evidence-based decision-making and root cause analysis instead of incident management.

Supervised Machine Learning Binary Classification

Three binary classifiers were used in predicting whether an incident was a breach or not. The model choice followed an evaluation criteria process where ROC AUC and recall were preferred to accuracy due to the practicality involved in the fact that there is more penalty for missing a breach compared to a false alarm.

Stakeholder Aligned Dashboard Design

The Power BI dashboard adheres to the principle of the layered intelligence design in that each page is set up according to the environment in which a particular stakeholder needs to make decisions. This practice is consistent with business analysis practices, where relevant information is provided to relevant people in the appropriate amount of detail. There will hardly be one consolidated dashboard suitable for everyone; there are three pages for three separate audiences.

As-Is / To-Be Process Analysis

The use of a structured analysis approach of “as is” and “to be” will enable a description of the current situation (volume based reactive management) and future situation (risk based proactive management).

3. Relevant Facts and Information

Key facts from the analysis:

- 24,918 total incidents sufficient for statistically reliable ML model training
- 9,115 SLA breaches 36.58% breach rate against industry benchmark of 85-90% compliance
- Category 46: 2,200 incidents (4th by volume), 1,114 breaches (1st by breach count) 51% implied breach rate
- Random Forest: 71.97% accuracy, 80.61% ROC AUC, 74.11% Recall, 65.92% F1 Score
- 391 incidents classified High Risk average breach probability 45.12%
- Top predictive features: subcategory_freq (12.46%), opened_by_freq (10.83%), u_symptom_freq (8.50%), open_hour (5.66%), priority_2 High (1.96%)
- Highest breach probability: INC0000378 and INC0000375 at 95.53%
- Financial exposure: 9,115 breaches x EUR 500 conservative benchmark = EUR 4.5M annual

Relationship to problem statement

The rate of breaches at 36.58% highlights the magnitude of this prioritization problem. The Category 46 anomaly proves this difference between the volume and the risks involved directly. The feature importance analysis shows that opened_by_freq and open_hour are some of the features that are completely not captured by the current ITSM reporting systems, which shows that the dashboards do not provide information needed to prevent breaches.

BUSINESS IMPACT

Current State vs Future State

Capability	Current State	Future State with Predictive Analytics
Incident Prioritisation	Volume based highest count categories get most attention	Risk-based Category 46 correctly identified as highest priority despite 4th place volume
SLA Management	Reactive breach reported after it occurs	Proactive 391 high risk incidents flagged before breach with probability scores
Resource Allocation	Uniform across shifts and teams	Time aware off peak hours identified as elevated risk; targeted staffing decisions

Management Reporting	Lagging KPIs breach rates from last period	Leading indicators predicted breach risk on current open incidents
Continuous Improvement	Experience based pattern recognition	Data-driven feature importance reveals specific subcategories and symptom types driving breach
Agent Performance	Subjective no data-driven visibility	Evidence-based opened_by_freq identifies teams associated with elevated breach rates

Financial Impact

Scenario	Basis	Estimated Value
Current breach exposure	9,115 breaches x EUR 500 benchmark penalty	EUR 4.5M annual exposure
10% breach reduction (conservative)	Proactive intervention on 391 high-risk incidents	EUR 450,000 annual saving
20% breach reduction (moderate)	Consistent with Cerence improvements using similar approach	EUR 900,000 annual saving
Reporting automation	3 hrs/week manual KPI x EUR 40/hr x 52 weeks	EUR 6,240 per analyst per year
MTTR improvement (15%)	Faster resolution through risk-based prioritisation	Reduced operational cost and improved client satisfaction

Transferability and Scalability

Application	Detail
Live ServiceNow integration	Python pipeline connects to ServiceNow REST APIs replacing synthetic data with production records
Category 46 equivalent identification	In any live environment the model identifies the hidden high risk category volume management overlooks
Agent performance analytics	Opened_by_freq creates evidence base for targeted training data-driven performance management
Shift pattern optimisation	open_hour importance enables data-driven staffing during highest

	breach risk hours
Scalability	Architecture scales from 50 person service desk to global 24x7 enterprise operations centre

CONCLUSION

Summary Of Findings

The research tried to solve one specific problem: why is there still a breach rate of 36.58% regarding SLAs in companies that have enough staff and proper ITSM processes? The result is clear and well-supported by numbers: the information required for avoiding the breaches is not visible to the people in charge of handling it.

Conclusion category 46 is what this conclusion is all about. There exists a category of incidents having a 51% violation rate which is the highest rate in the whole set of data which was not visible until the analysis was done. When resource allocation was made for Category 42, it was a decision that seemed rational considering the amount of information they had. The problem here is that the information was wrong.

Random Forest achieves an ROC AUC score of 80.61%, which is able to successfully distinguish between breach and non-breach on an incident by incident basis. The top predictors for this mode subcategory_freq, opened_by_freq, u_symptom_freq, and open_hour are all factors that are currently being measured in any ITSM system. This intelligence has always been there but never leveraged before.

The financial case is clear cut; the one-off investment for analysis provides a yearly reduction in breaches of between EUR 450,000 and EUR 900,000, along with EUR 6,240 savings per analyst in reporting efficiencies, which is significantly less expensive than just one large-scale event.

Implications of the Study

Implications go far beyond the current dataset. In all the companies that provide ITSM (Information Technology Service Management) services on a large scale, the same gap in information is faced. The data processing pipeline implemented in Python, model structure, and dashboard design discussed above can be easily implemented in live ServiceNow or Jira systems. There will always be Category 46 in any functioning ITSM system, and the model can identify it after several days of its implementation.

The agency level finding provides the evidential basis required for workforce development activities that subjective based performance management does not provide. The time of day finding facilitates planning for shifts using breach probability information rather than conventional wisdom. The three page dashboard structure enables strategic, tactical, and analytic information delivery to three different target groups through one source of data.

REFERENCES

Source	Detail	Link
Primary Dataset	UCI Machine Learning Repository ITSM dataset used as basis for synthetic data generation and feature modelling	archive.ics.uci.edu
ML Framework	Scikit-Learn Random Forest, Logistic Regression, Gradient Boosting classifiers	scikit-learn.org
Visualisation Tool	Microsoft Power BI with DAX measures	powerbi.microsoft.com
ITIL Framework	ITIL v4 Foundation Continual Service Improvement, Problem Management, Service Level Management	axelos.com
Industry Benchmark	HDI (Help Desk Institute) Service Management SLA compliance benchmarks	thinkhdi.com
GitHub Repository	ITSM Predictive Analytics Dashboard full code, data and documentation	github.com/kartikkodilkar-netizen/ITSM-Predictive-Analytics-Dashboard
Live Dashboard	Interactive Power BI dashboard Projects section	kartikkodilkar.org
Operational Experience	7+ years enterprise ITSM TCS (2017-2021), Cerence (2021-2022), Accenture Dublin (2025-present)	kartikkodilkar.org

Author	Kartik	kartikkodilkar.org
Contact	Kodilkar|kartik.kodilkar@gmail.com| +353 83 854 7122	