ITSM Machine Learning Project

Project Report

|  |  |
| --- | --- |
| Project Name | ITSM Machine Learning Project |
| Project Ref | PM-PR-0012 |
| Document Ref | PR0012-Project-Report |
| Version | 1.2 |
| Client | ABC Tech |
| Description | Machine Learning to improve incident management process |
| Project Sponsor | Leon Van der vault, Director - ABC Tech |
| Author(s) | Ranjith, Project Manager, Data Scientist |
| Last Updated Date | Jan 10, 2019 |

# Contents

[Contents 2](#_Toc535325872)

[1 Project Summary 3](#_Toc535325873)

[2 Problem Definition 3](#_Toc535325874)

[3 Mapping to Machine Learning 3](#_Toc535325875)

[1 The target function / Variable 4](#_Toc535325876)

[2 Feature representation 4](#_Toc535325877)

[3 System structure 4](#_Toc535325878)

[4 Data Preparation 4](#_Toc535325879)

[5 EDA – Exploratory Data Analysis 4](#_Toc535325880)

[6 Learning System Design 4](#_Toc535325881)

[1 The learning algorithm 4](#_Toc535325882)

[2 Improvements/modifications 5](#_Toc535325883)

[7 Model Evaluation / Testing Results 5](#_Toc535325884)

[1 The data sets 5](#_Toc535325885)

[2 Learning 5](#_Toc535325886)

[3 Results & Data analysis 5](#_Toc535325887)

[8 Retraining and Future Work 6](#_Toc535325888)

[9 Model Deployment 6](#_Toc535325889)

[10 Project Performance 6](#_Toc535325890)

[1 Schedule 6](#_Toc535325891)

[2 Budget 7](#_Toc535325892)

[11 Lessons Learned 7](#_Toc535325893)

[12 Conclusion 7](#_Toc535325894)

# Project Summary

ABC Tech is a mid-size organisation operation in IT-enabled business segment over a decade. On an average ABC Tech receives 22-25k IT incidents/tickets, which were handled to best practice ITIL framework with incident management, problem management, change management and configuration management processes. These ITIL practices attained matured process level and a recent audit confirmed that further improvement initiatives may not yield return of investment.

ABC Tech management is looking for ways to improve the incident management process as recent customer survey results shows that incident management is rated as poor.

ABC Tech management recently attended Machine Learning conference on ML for ITSM.

Machine learning looks prospective to improve ITSM processes through prediction and automation. They came up with 4 key areas, where ML can help ITSM process in ABC Tech.

1. Predicting High Priority Tickets: To predict priority 1 & 2 tickets, so that they can take preventive measures or fix the problem before it surfaces.
2. Forecast the incident volume in different fields, quarterly and annual. So that they can be better prepared with resources and technology planning.
3. Auto tag the tickets with right priorities and right departments so that reassigning and related delay can be reduced.
4. Predict RFC (Request for change) and possible failure / misconfiguration of ITSM assets.

This report is for Key area 1:

Predicting High Priority Tickets: To predict priority 1 & 2 tickets, so that they can take preventive measures or fix the problem before it surfaces.

# Problem Definition

1. Recent customer survey results of ABC Tech show that incident management is rated as poor in spite of ABC Tech implement ITIL best practice process.
2. Time taken to identify and assign High Priority ticket often the main reason for breach in Service Level Agreement of not resolving in time.

# Mapping to Machine Learning

## features and data description

Data Description

CI\_Name 46606 non-null object

CI\_Cat 46495 non-null object

CI\_Subcat 46495 non-null object

WBS 46606 non-null object

Incident\_ID 46606 non-null object

Status 46606 non-null object

Impact 46606 non-null object

Urgency 46606 non-null object

Priority 45226 non-null float64

Category 46606 non-null object

KB\_number 46606 non-null object

Alert\_Status 46606 non-null object

No\_of\_Reassignments 46605 non-null float64

Unnamed: 14 46606 non-null int64

Open\_Time 46606 non-null datetime64[ns]

Reopen\_Time 2284 non-null datetime64[ns]

Resolved\_Time 44826 non-null datetime64[ns]

Close\_Time 46606 non-null datetime64[ns]

Handle\_Time\_hrs 46605 non-null object

Closure\_Code 46146 non-null object

No\_of\_Related\_Interactions 46492 non-null float64

Related\_Interaction 46606 non-null object

No\_of\_Related\_Incidents 1222 non-null float64

No\_of\_Related\_Changes 560 non-null float64

Related\_Change 560 non-null object

## The target function / Variable

Target Variable: Priority

## Feature representation

The features are selected based on the correlation as well as trial and error basis to get best model fitment.

CI\_Subcat

WBS

Category

# Data Preparation

Data contained 46606 rows of data with 26 colouns

RangeIndex: 46606 entries, 0 to 46605

Data columns (total 26 columns):

CI\_Name 46606 non-null object

CI\_Cat 46495 non-null object

CI\_Subcat 46495 non-null object

WBS 46606 non-null object

Incident\_ID 46606 non-null object

Status 46606 non-null object

Impact 46606 non-null object

Urgency 46606 non-null object

Priority 45226 non-null float64

Unnamed: 9 46606 non-null float64

Category 46606 non-null object

KB\_number 46606 non-null object

Alert\_Status 46606 non-null object

No\_of\_Reassignments 46605 non-null float64

Unnamed: 14 46606 non-null int64

Open\_Time 46606 non-null datetime64[ns]

Reopen\_Time 2284 non-null datetime64[ns]

Resolved\_Time 44826 non-null datetime64[ns]

Close\_Time 46606 non-null datetime64[ns]

Handle\_Time\_hrs 46605 non-null object

Closure\_Code 46146 non-null object

No\_of\_Related\_Interactions 46492 non-null float64

Related\_Interaction 46606 non-null object

No\_of\_Related\_Incidents 1222 non-null float64

No\_of\_Related\_Changes 560 non-null float64

Related\_Change 560 non-null object

dtypes: datetime64[ns](4), float64(6), int64(1), object(15)

Only two columns were relevant (CI\_sub, WBS, Category) for the model.

As the missing value records are less than 5%, they were dropped resulting 45115 records.

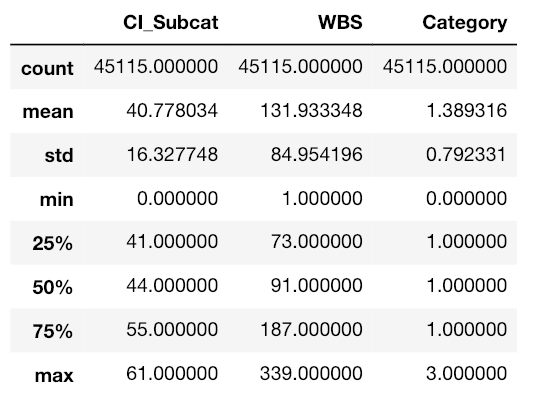
CI\_Subcat 45118 non-null object

WBS 45118 non-null object

Category 45118 non-null object

# EDA – Exploratory Data Analysis

Selected columns descriptive stats.



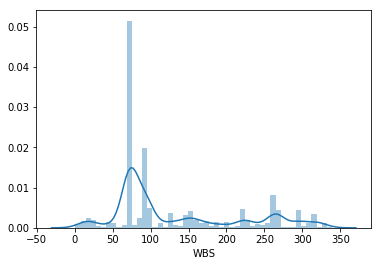
Ticket Priority Volumes in data set:

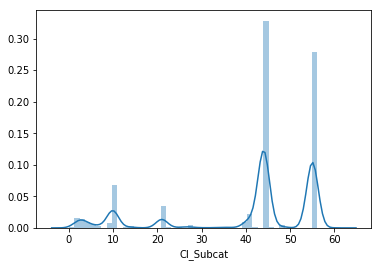
2: 689

3: 5310

4: 22689

5: 16427





# Model Evaluation / Testing Results

## The data sets

Train Data Set size: 33836

Test Data Set size: 11279

## Learning

Evaluated for Three Learning models with Random Forest, XgBoost and SVM.

Random Forest and SVM Classifier yielded similar results with 83.6% accuracy.

SVM Classifier with parameters below got best fitment

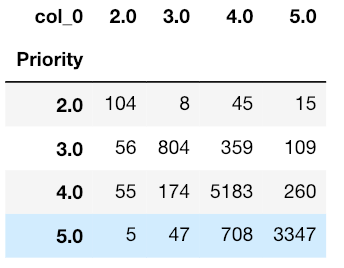
Kernel: rbf

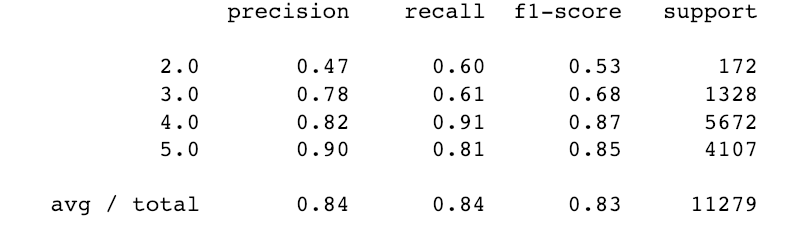
C = 10

Gamma = 0.1

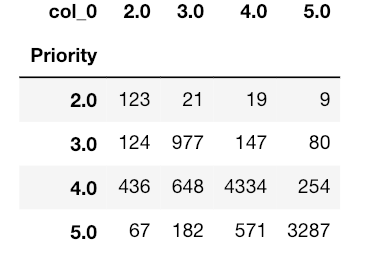
## Results & Data analysis

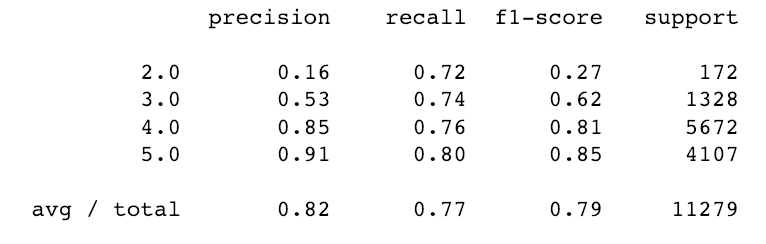
Confusion Matrix with SVM without balancing the dataset.





Confusion Matrix with SVM with balancing the dataset with SMOTE technique.





Balancing data with SMOTE increase recall rate of priority 2 and 3 from 60% and 61% to 72% and 74% respectively.

As the recall rate is more important in this business case, which is expected to find high priority tickets, the second results with SMOTE is considered as best and used in final model.

# Re-training and Future Work

The data given already has human bias and not correctly updated fields which is effecting model efficiency. The future work may collect the incident data with expert focus group for the purpose of machine learning .

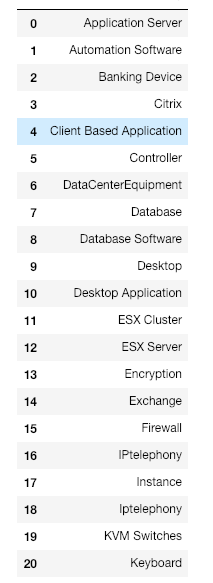
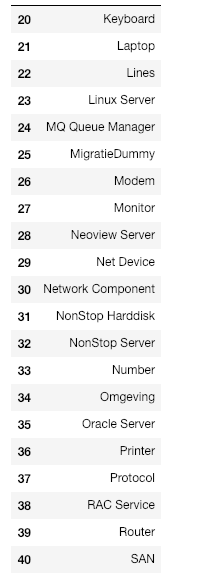
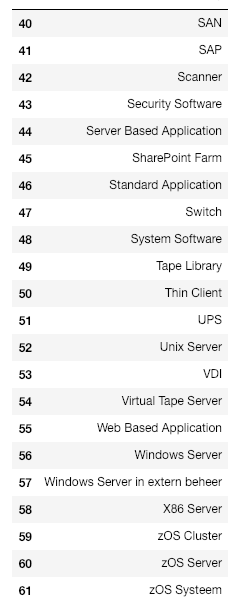
This can be used for training and thereby resulting model can achieve high efficiency, model precision and recall rates.

# Model Deployment

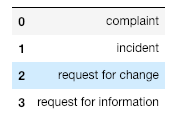
Model takes features : CI\_Subcat, WBS and Category as input and returns priority range [2, 3,4,5]

The features were label encoded for model training. The labels were as given below. The deployment of this model should follow the same labelling convention.

### CI\_Subcat

### Category



### WBS

Last integer part is taken

For example 'WBS000263 is taken as 263

# Project Performance

The designed model give about 75% recall rate means that 25% of the highest priority tickets are missed. Though this does improve the work follow, a recall rate of 85% would have been significant value addition.

## Schedule

Top of Form

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *On Schedule* | | | | | |
| Planned Finish Date | Actual Finish Date | Variance  (in days) | On Schedule | Ahead of Schedule | Behind Schedule |
| 15 Jan 2019 | 14 Jan 2019 | -1 |  |  |  |
| *\*’On Schedule’ calculation may be within +/- 10% of the Approved Schedule* Budget | | | | | |
| *On Budget* | | | | | |
| Approved Budget | Spent Budget | Variance  (in $) | On Budget | Under Budget | Over Budget |
| 100k | 100k | 0 |  |  |  |
| *\*’On Budget’ calculation may be within +/- 10% of the Approved Budget* | | | | | |
| *Meeting Customer Expectations* | | | | | |
| Success Criteria | | | Criteria Met | Comments | |
| 80+ recall rate in Priority 2 and 3 | | |  | Not meet but model is value to business as it got 75% recall rate | |

Bottom of Form

# Conclusion

ITSM Machine Learning project partially met the project objectives of predicting high priority tickets. The model can be deployed with a recall rate of 75% (approx.) for Priority 2 and 3 tickets.

Priority 1 tickets are only a few and not enough for ML modeling so excluded for the data set.