Northwestern University IT Ticket Times
Rene Romo (rgr355),Ryan Miller(rsm408)

The Problem:
For this project our goal is to be able to predict the length of time it takes Northwestern IT to get an issue resolved, from creation to resolution, depending on attributes such as the issue's category, when it was created, what department had to handle it, its urgency etc.. The reason we have chosen this task is that we have worked at NUIT and very often people will ask for an estimate of how long an issue will take to resolve. Currently there is not an effective way to predict how long this will take and so we have to either tell the user that we can't provide an estimate or provide our best and risk upsetting the user if it is not resolved in that time. We have also noticed that there are trends for how long an issue takes based on the type of issue,who is handling it, when it was submitted, etc. A proper use of Machine Learning could allow us to provide more accurate estimates to users and identify any slowdowns within departments or certain types of issues that don't make sense and suggest a department may be understaffed or that the solution to a problem is inefficient.

The Data:
We collected data using the FootPrints Ticketing System's reports feature. This allowed us to export data regarding all the support requests received in the month of April 2016. For security and privacy reasons we removed any requests involving our security team and used only attributes that would not be personally identifiable. When we gathered the data we gathered the following attributes:

- **Date Submitted -** Initially in the format MM/DD/YYYY
- **Time Submitted -** Initially in the format HH:MM:SS
- **Last Edit Date**
- **Last Edit Time**
- **Assignees-** This is the group who handled and resolved the issue
- **Service Family/Service/Category-** What services the issue involved (e.g. Canvas Issue or Wifi Configuration etc.
- **Urgency-** Routine,Elevated, Critical, or Immediate
- **Department/Departmental Support** - the department of the person who submitted the issue (eg. McCormick Mechanical Engineering, Feinberg Neurology etc)
- **NU Role -** Faculty/Staff, Student, Temp, or Affiliate
- **VIP Status -** DDCA(Dean, Department Chair or Administrator) or None
- **Time To Close -** Time from creation to close In Minutes

Our Final Data Set consisted of ~9700 total tickets. Of these we pulled out 1000 randomly sampled tickets to be our test set. ended up having to remove tickets which had special characters that were causing trouble when we attempted to create arff files for Weka. Our final training set included 8690 instances and the test set had 983.

Throughout the project we found that some attributes should be reworked in order to provide the most useful amount of information. First and most importantly we decided rather than have a Numerical output in Time To Close in number of minutes, we should instead bin this into a nominal set of values as for our purposes of providing an estimate of how long an issue would take this was more than sufficient and would lead to better accuracy. We went back and forth but ultimately decided we would split the Time to Close category into bins of 12 hour increments with any issues taking over 5 days(120 hours) being placed in the same bin. (e.g. Our bins were: <12 hours, 12-24, 24-36….120+). We found that the Last Edit Time and Date Fields were not useful for our machine learning task. Additionally we found that the attributes of time and date submitted would be best replaced by Nominal Attributes Indicating whether the issue was submitted on a Weekday or Weekend and during Business Hours or Non-business hours.

Figure 1: Typical Instance

Date Submitted	Time Submitted	Assignees	Service Family	Service	Category	Urgency	Department	Departmental	NU Role	VIP Status	Output
Weekday	Non-Business Hours	NUIT-TSS-SupportServices:	Hardware & Software	Other		Routine	IT Technology Support Svcs	NUIT-DSS	staff		120+ hours

Methodology:

We decided to use Weka for our project in order to try a wide range of algorithms and focus on tuning parameters and trying to achieve the best accuracy with our given data. Because we wanted to be able to identify any choke points within NUIT we decided to focus on the ‘rules’ category of Weka classifiers so that we could see which attributes were most important to determining the time it took to resolve an issue and whether there was any noticeable anomalies. Notably we tried Decision Tables,Ripple Down Rules, and REPTrees.We also tested classifiers covered in class such as Naive Bayes and K-nearest nearest classifiers. Ultimately we found the the two most accurate algorithms for our data were the J48 decision tree algorithm and IBk Nearest Neighbor Algorithms

Results:

Our results are summarized in the tables below. The first is the accuracy versus our test set of 983 instances. We can see that while there is some improvement as to just always guessing the majority category of “Less than 12 hours”, it is not a vast improvement and due to the small amount of test data cannot be said to be of any statistical significance due to high p-values. The second table is the results of 10 fold cross validation across the 8690 instances in our training set. The improvement over Zero R here are extremely statistically significant with p-values less than 0.0001

These results were generated with the following tuned configurations

J48:

weka.classifiers.trees.J48 -C 0.25 -B -M 10 -A

IBk:

weka.classifiers.lazy.IBk -K 8 -W 0 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\""

Test Set:

Algorithm	Correct	Incorrect	Accuracy
Zero R	613	370	62.3%
J48	654	329	66.5%
IBk k=8	652	331	66.3%

J48 654/329, The two-tailed P value equals 0.0594
IBk k=8 652/331, The two-tailed P value equals 0.0735

Ten Fold Cross Validation with Training Set:

Algorithm	Correct	Incorrect	Accuracy
Zero R	5505	3185	63.3%
J48	5834	2856	67.1%
IBk k=8	5857	2833	67.3%

J48: 5834/2856, The two-tailed P value is less than 0.0001
IBk k=8 5857/2833, The two-tailed P value is less than 0.0001

Using the pruned decision tree generated by the J48 algorithm(Appendix A) we can see that the most important attributes are, whether an issue was assigned to the main Support Center, whether it was submitted on a weekday or weekend, and it’s assignment group. Additionally we were able to identify some possibly problematic areas such as issues having to do with Our Northwestern being more likely to take more than five days which could be further investigated to improve efficiency.

Future Work:

While in the end our project did not lead to as accurate results as we'd originally hoped we still believe that with further work machine learning could be used to accurately predict the time it will take to resolve an issue. For future work we recommend using more data, we found that even with all the tickets for a month more than half fell within the category of less than 12 hours which we believe skewed our results, especially when using K-Nearest Neighbors. More data would allow us to garner more information and differentiate between the categories, Additionally we believe there are more attributes that we did not use which may have more of an impact than originally thought such as who was working on it, what support tools were used,whether the request was made in person, as a call or in an online chat, and attributes relating to the text from the title and description of the request. With these features we believe that the accuracy could be significantly improved.

Who did what:

The project was done entirely through group meetings with the exception of the Final Report and Website which we split up each taking one. Website-Ryan, Report - Rene.

Appendix A:

```
Assignees = A_IT SUPPORT CENTER:: Less than 12hrs (2507.0/58.0)
Assignees != A_IT SUPPORT CENTER:
|  Date Submitted = Weekday
|  |  Service = Our Northwestern
|  |  |  Category = Alumni Info: 120+ hours (190.46/39.33)
|  |  |  Category != Alumni Info
|  |  |  |  Assignees = NUIT-ADEA-OurNorthwestern:
|  |  |  |  |  Category = Other: Less than 12hrs (13.06/5.04)
|  |  |  |  |  Category != Other: 120+ hours (102.49/49.45)
|  |  |  |  Assignees != NUIT-ADEA-OurNorthwestern:: Less than 12hrs (104.2/48.43)
|  |  Service != Our Northwestern
|  |  |  Time Submitted = Non-Business Hours
|  |  |  |  Assignees = NUIT-FFRA-Technical:: 24-36 hours (10.0/7.0)
|  |  |  |  Assignees != NUIT-FFRA-Technical:
|  |  |  |  |  Department = Preventive Medicine; Feinberg School of Medicine: 12-24 hours (35.9/23.41)
|  |  |  |  |  Department != Preventive Medicine; Feinberg School of Medicine
|  |  |  |  |  |  Department = IT Cyber Service Operations: 120+ hours (30.95/16.0)
|  |  |  |  |  |  Department != IT Cyber Service Operations
|  |  |  |  |  |  |  Service = Desktop/Laptop/Smartphone/Tablet: 120+ hours (46.03/29.19)
|  |  |  |  |  |  |  Service != Desktop/Laptop/Smartphone/Tablet
|  |  |  |  |  |  |  |  Assignees = Feinberg-IT-Tier-2:: Less than 12hrs (23.76/8.88)
|  |  |  |  |  |  |  |  Assignees != Feinberg-IT-Tier-2:
|  |  |  |  |  |  |  |  |  Department = IT Technology Support Svcs: 120+ hours (51.78/31.27)
|  |  |  |  |  |  |  |  |  Department != IT Technology Support Svcs
|  |  |  |  |  |  |  |  |  |  Service = Client Backup (CrashPlan): Less than 12hrs (14.14/5.19)
|  |  |  |  |  |  |  |  |  |  Service != Client Backup (CrashPlan)
|  |  |  |  |  |  |  |  |  |  |  Assignees = NUIT-ADEA-CATracksReports:: 120+ hours (16.0/9.0)
|  |  |  |  |  |  |  |  |  |  |  Assignees != NUIT-ADEA-CATracksReports:
|  |  |  |  |  |  |  |  |  |  |  |  Assignees = NUIT-ART-LSTS:: Less than 12hrs (29.62/14.85)
|  |  |  |  |  |  |  |  |  |  |  |  Assignees != NUIT-ART-LSTS:: 12-24 hours (306.28/195.44)
|  |  |  Time Submitted != Non-Business Hours
|  |  |  |  Category = Deployment/Setup
|  |  |  |  |  Assignees = Feinberg-IT-Onboarding:: 120+ hours (59.62/20.23)
|  |  |  |  |  Assignees != Feinberg-IT-Onboarding:
|  |  |  |  |  |  Assignees = Feinberg-IT-Deployment:: 120+ hours (72.31/30.93)
|  |  |  |  |  |  Assignees != Feinberg-IT-Deployment:
|  |  |  |  |  |  |  Service = Desktop/Laptop/Smartphone/Tablet
|  |  |  |  |  |  |  |  Assignees = Feinberg-IT-Tier-1:: Less than 12hrs (20.0/11.0)
|  |  |  |  |  |  |  |  Assignees != Feinberg-IT-Tier-1:
|  |  |  |  |  |  |  |  |  NU Role  = faculty : Less than 12hrs (10.2/5.14)
|  |  |  |  |  |  |  |  |  NU Role  != faculty : 120+ hours (93.11/58.11)
|  |  |  |  |  |  |  Service != Desktop/Laptop/Smartphone/Tablet: Less than 12hrs (134.47/58.65)
|  |  |  |  Category != Deployment/Setup
```

| | | | | Assignees = UniversityEnrollment-RO-Security: : 120+ hours (16.0/5.0)
| | | | | Assignees != UniversityEnrollment-RO-Security:
| | | | | | Assignees = Feinberg-IT-Onboarding:: 120+ hours (13.38/4.77)
| | | | | | Assignees != Feinberg-IT-Onboarding:
| | | | | | | Service Family = Security and Compliance
| | | | | | | | Service = Other: 72-84 hours (16.72/8.41)
| | | | | | | | Service != Other: Less than 12hrs (13.93/7.41)
| | | | | | | Service Family != Security and Compliance
| | | | | | | | Category = Ad-Hoc Reports
| | | | | | | | | Assignees = NUIT-ADEA-CATracksReports:: 120+ hours (39.0/12.0)
| | | | | | | | | Assignees != NUIT-ADEA-CATracksReports:: Less than 12hrs (22.51/9.36)
| | | | | | | | Category != Ad-Hoc Reports
| | | | | | | | | Service = IT Service Manager: 120+ hours (18.13/13.12)
| | | | | | | | | Service != IT Service Manager
| | | | | | | | | | Service = NUInfo/Central Web Server: 120+ hours (11.0/6.44)
| | | | | | | | | | Service != NUInfo/Central Web Server
| | | | | | | | | | | Service = Server
| | | | | | | | | | | | Category = Virtual Server Request: 120+ hours (16.11/10.11)
| | | | | | | | | | | | Category != Virtual Server Request: Less than 12hrs (38.28/19.12)
| | | | | | | | | | | Service != Server
| | | | | | | | | | | | Service = AutoSys Job Scheduler
| | | | | | | | | | | | | Assignees = NUIT-CI-SO-QualityControl:: Less than 12hrs (15.0/9.0)
| | | | | | | | | | | | | Assignees != NUIT-CI-SO-QualityControl:: 120+ hours (10.18/5.16)
| | | | | | | | | | | | Service != AutoSys Job Scheduler
| | | | | | | | | | | | | Assignees = NUIT-CI-TNS-SSLVPN:: 12-24 hours (10.91/4.91)
| | | | | | | | | | | | | Assignees != NUIT-CI-TNS-SSLVPN:
| | | | | | | | | | | | | | Assignees = Feinberg-IT-ClientManagement:: 120+ hours (19.47/12.1)
| | | | | | | | | | | | | | Assignees != Feinberg-IT-ClientManagement:
| | | | | | | | | | | | | | | Assignees = NUIT-ART-ResearchComputing:
| | | | | | | | | | | | | | | | Service = Social Sciences Computing Cluster (SSCC): Less than 12hrs (16.57/9.29)
| | | | | | | | | | | | | | | | Service != Social Sciences Computing Cluster (SSCC): 120+ hours (33.01/15.55)
| | | | | | | | | | | | | | | Assignees != NUIT-ART-ResearchComputing:
| | | | | | | | | | | | | | | | Category = Daily Application Audit Report: 120+ hours (15.47/7.76)
| | | | | | | | | | | | | | | | Category != Daily Application Audit Report: Less than 12hrs (4322.93/1758.21)
| Date Submitted != Weekday
| | Assignees = NUIT-ART-ResearchComputing:: Less than 12hrs (16.0/7.0)
| | Assignees != NUIT-ART-ResearchComputing:
| | | Time Submitted = Non-Business Hours
| | | | NU Role = staff: 120+ hours (24.41/9.55)
| | | | NU Role != staff
| | | | | NU Role = staff : 120+ hours (30.22/11.16)
| | | | | NU Role != staff : 12-24 hours (31.38/24.01)
| | | Time Submitted != Non-Business Hours
| | | | Assignees = OARD-Our Northwestern:: 12-24 hours (11.0/5.0)
| | | | Assignees != OARD-Our Northwestern:
| | | | | NU Role = staff: 120+ hours (16.76/11.47)
| | | | | NU Role != staff
| | | | | | Service Family = Email, Calendaring, Messaging and Collaboration: 12-24 hours (12.33/6.63)
| | | | | | Service Family != Email, Calendaring, Messaging and Collaboration: 36-48 hours (27.9/17.9)

Number of Leaves : 47

Size of the tree : 93