Predicting Network Fault Severity

Capstone Presentation

by Ivy Chiang

Introduction

- Background
  - Data Set
  - Exploratory Analysis
  - Model Building
  - Results and Discussion
- Recommendation
  - Further studies

Background

: Predicting Network Fault Severity :

Why?

Customers

- Emergencies
- Communication
- Business
- Information

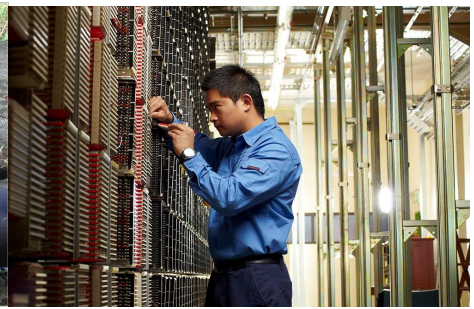
Telecom Company

- Service output
- Maintenance cost
- Reputation
- Compensation

Background

: Reasons - service disruption :



High service volume
- Emergencies
- Celebrations

Physical damage
- Storms

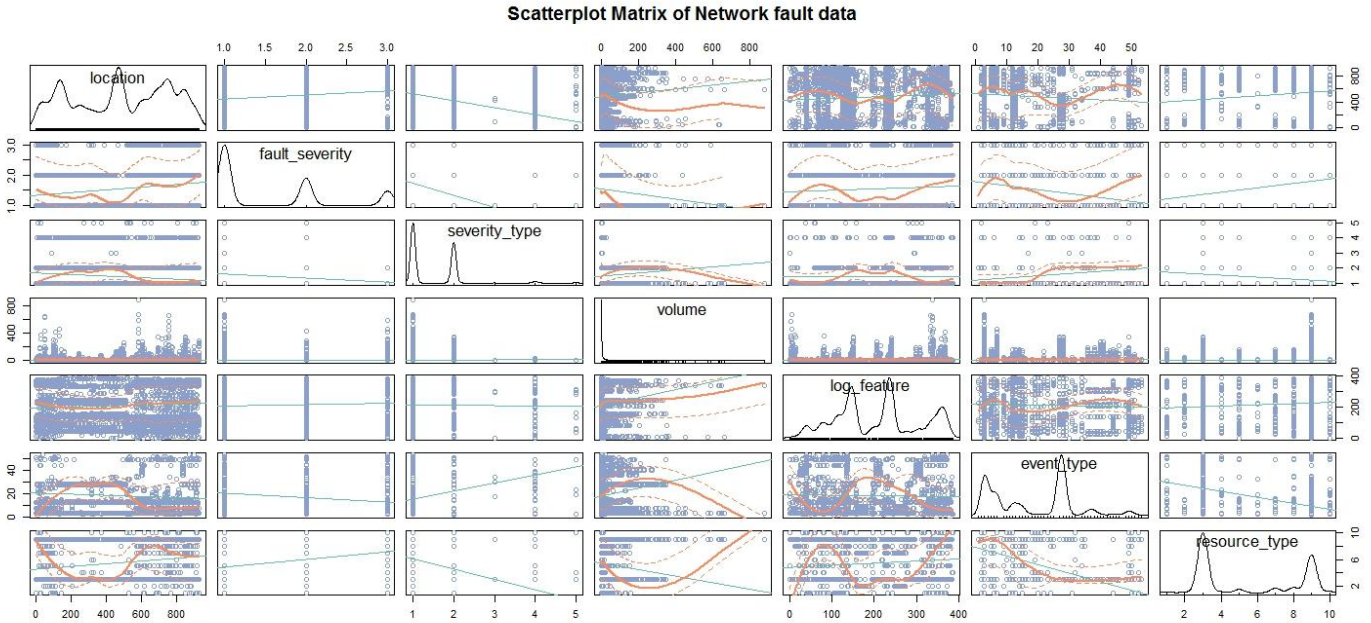
Repair Maintenance
- Known
- unknown

Dataset

Data name	Data type	Brief description
id	integer	Record's identifier , time point
location	Integer (factor)	Fault location
faulty_severity	Integer (factor)	3 levels, 0= no fault ; 1 = few ; 2 = many - actual data from reported fault from users
log_feature	Integer (factor)	Assumption: network service's faulting feature
volume	integer	Assumption: unit measure of faulting feature
resource_type	Integer (factor)	10 distinct types
event_type	Integer (factor)	Assumption: Network service feature's fault behaviour
severity_type	Integer (factor)	5 unordered types - warning message received from monitoring machines

- **Data Cleaning** - type conversion, trimming, splitting
- **Missing data** - 2 tables - network & incident_log
- **Limitation** - assumption of feature = output feature , event = fault behaviour

Exploratory Analysis

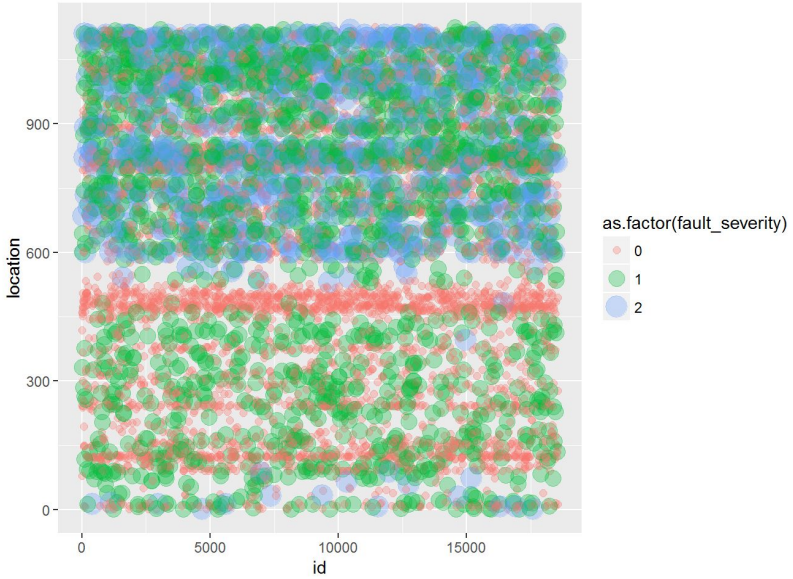


Scatterplot matrix - discrete distribution, mapping patterns and cross-relations

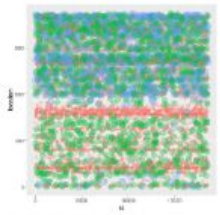
Exploratory Analysis

Relationship
features by id (time)

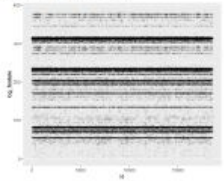
- Some locations
- always on



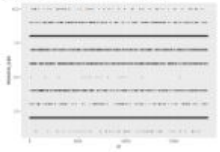
(1) Location



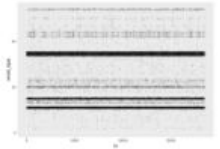
(2) Log_feature



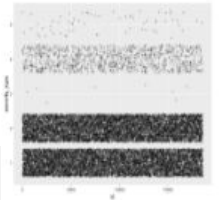
(3) Resource_type



(4) Event_type



(5) Severity_type



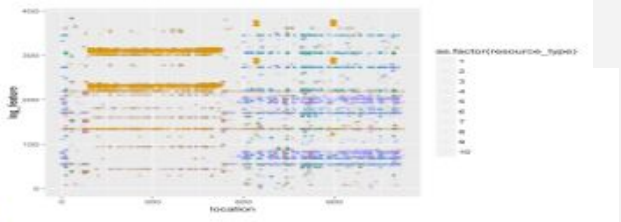
Exploratory Analysis

Feature start to separate

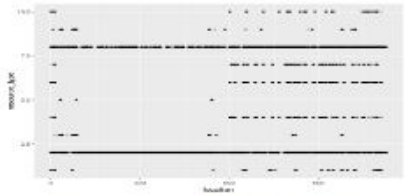
Relationship **log feature by location**

- Micmicking Resource_type
- Event_type
- Severity_type

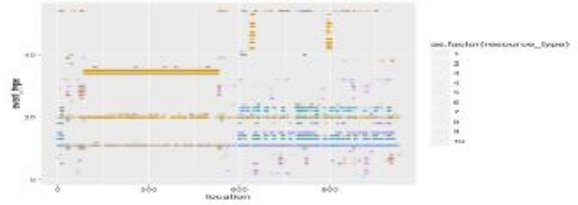
(1a) col = resource



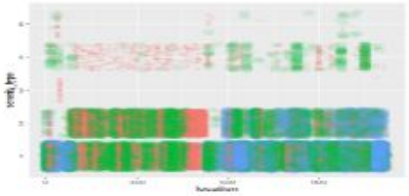
(3) Resource_type



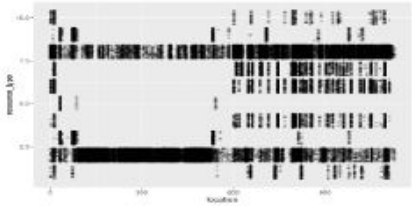
(4) Event_type



(5) Severity_type



(6) Resource: Jitter

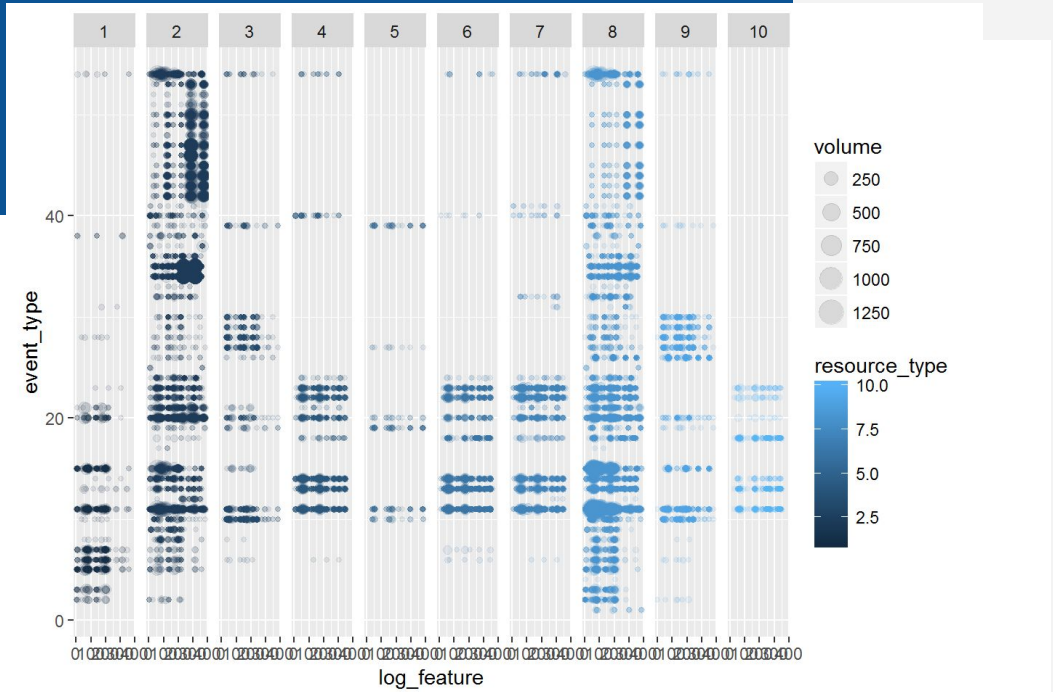


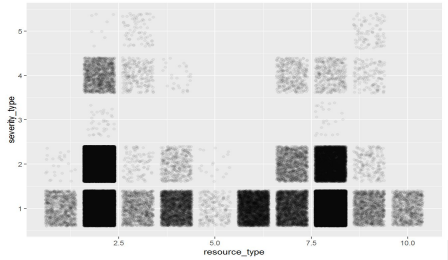
Exploratory Analysis

Distinct **event_types** to describe each **resource_type**

Set pattern between **event_type** and **Log_feature**

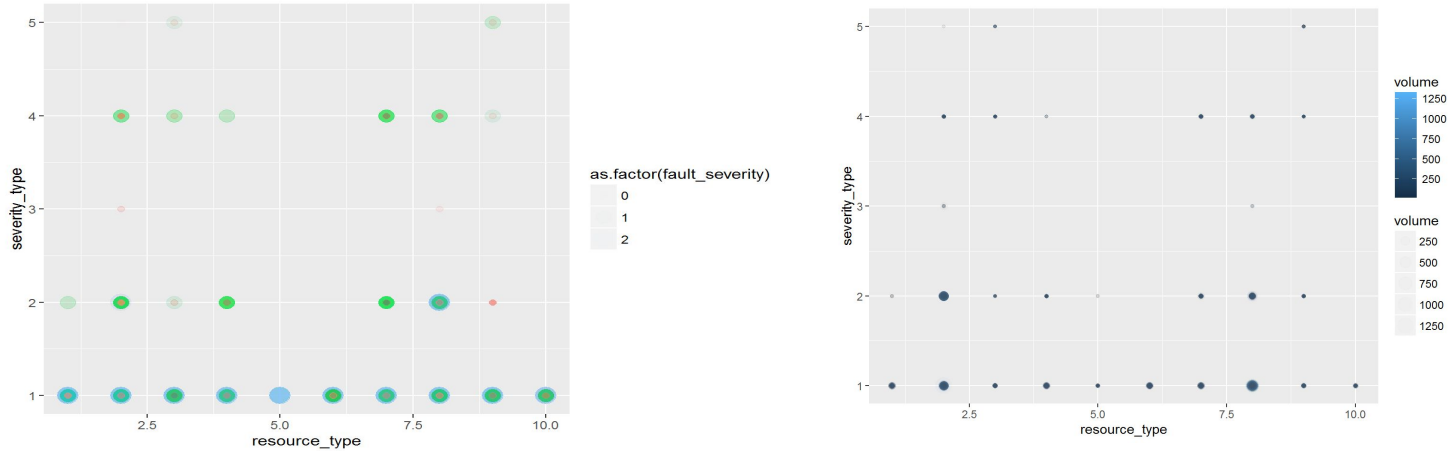
Which also correlates with **severity_type** X **resource_type**





Exploratory Analysis

Then talking about **Severity_type** & **resource_type**,
there's a ***similarity*** with scales of **fault_severity** & **volume**



Model Building

Random Forest - for categorical, complex non-linear data
- use of decision trees

Feature Engineering - simplify

1. Combination study
2. Extract counts

Model Building

Features

variable	description	reason
resource_type	unit measure of faulting feature	High correlation in scatterplot
severity_type	5 unordered types - warning message received from monitoring machines	High correlation in scatterplot
volume	unit measure of faulting feature	High correlation in scatterplot
t_feature	Tally of unique service that location registers fault on	Influences portion of logs against the location
t_resource	Tally of unique resources that location utilises	Influences portion of logs against the location
t_event	Tally of unique events occurring at that location	High relation to feature and severity_type
sev_res.rf	Probability of severity_type rating with a particular	Gives a weighing

Results and Discussion

Winning model:

Fault_severity ~
t_feature + sev_res.rf + volume + t_event

Overall Accuracy ~ 74%

Results and Discussion

To predict information need :

1. No# of feature services the location performs - ***How busy?***
2. The ***knowledge*** of the resource_type's **severity message type**?
3. The service volume output - ***How busy?***
4. Number of frequently logged events - ***Fault history range***

Results and Discussion

Limitation :

1. Other location specific variable - metro / climate
2. Accuracy of the severity_type message
3. Locations of low traffic hence **fault** -> no user impact
4. ...Or Excellent disruption bypass routes

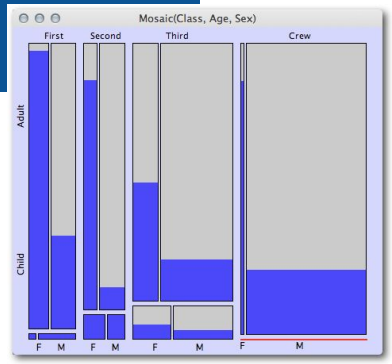
Recommendation

1. Unique alert from high traffic service locations.
   currently fault_severity rating 2 is clouded by type 1 or 2 band

2. Invest innovative monitoring system to ensure accuracy severity_type warnings

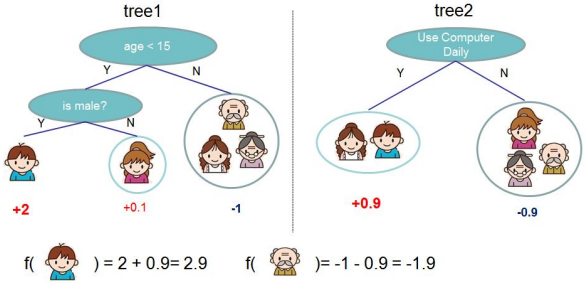
3. Quality materials and workmanship to prolong equipment life. Avoid aging equipment that attract long-term accumulated maintenance cost

Further studies



1. Better feature engineering - max use of data
   a. Probability theories
   b. Mosaic plotting

2. Hybrid use of modelling methods with Random Forest
   a. Clustering
   b. Gradient boosting (XGBoost)



thank you !

Introduction

sample