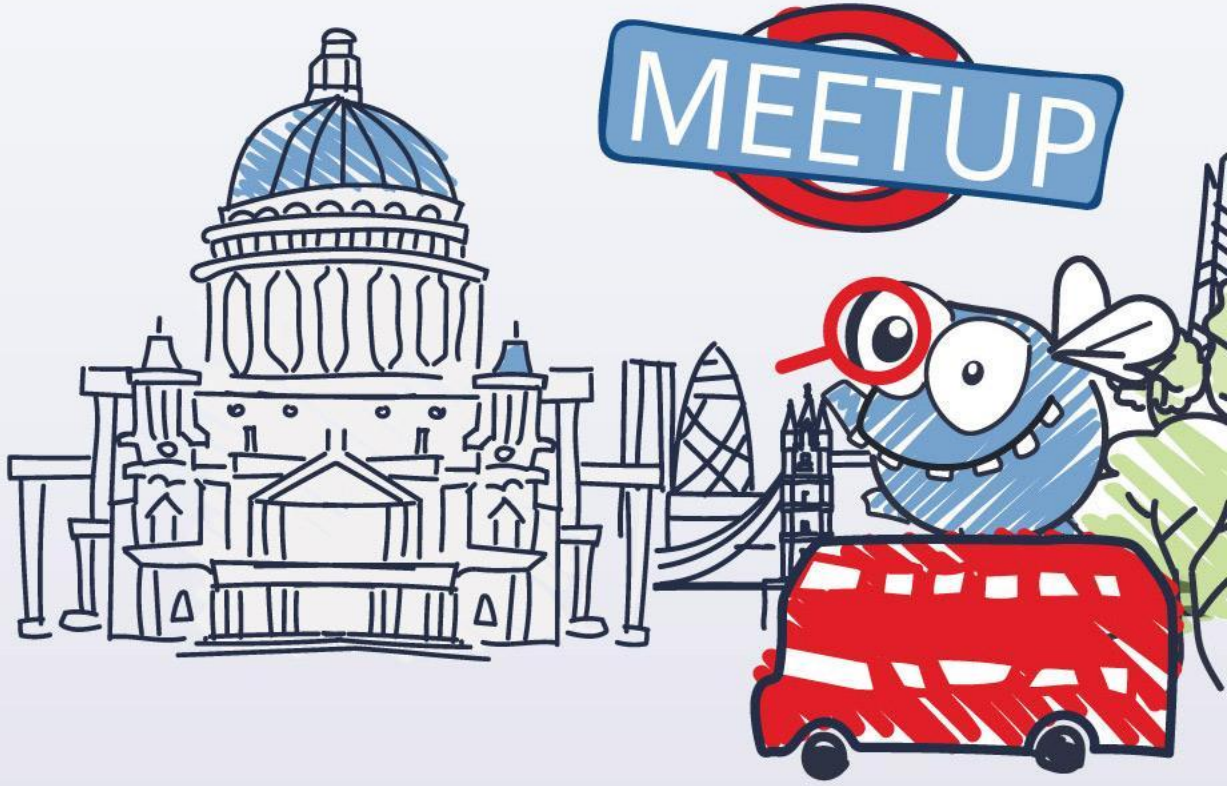
EXTENT CONFERENCE

Machine learning applied to defect report analysis



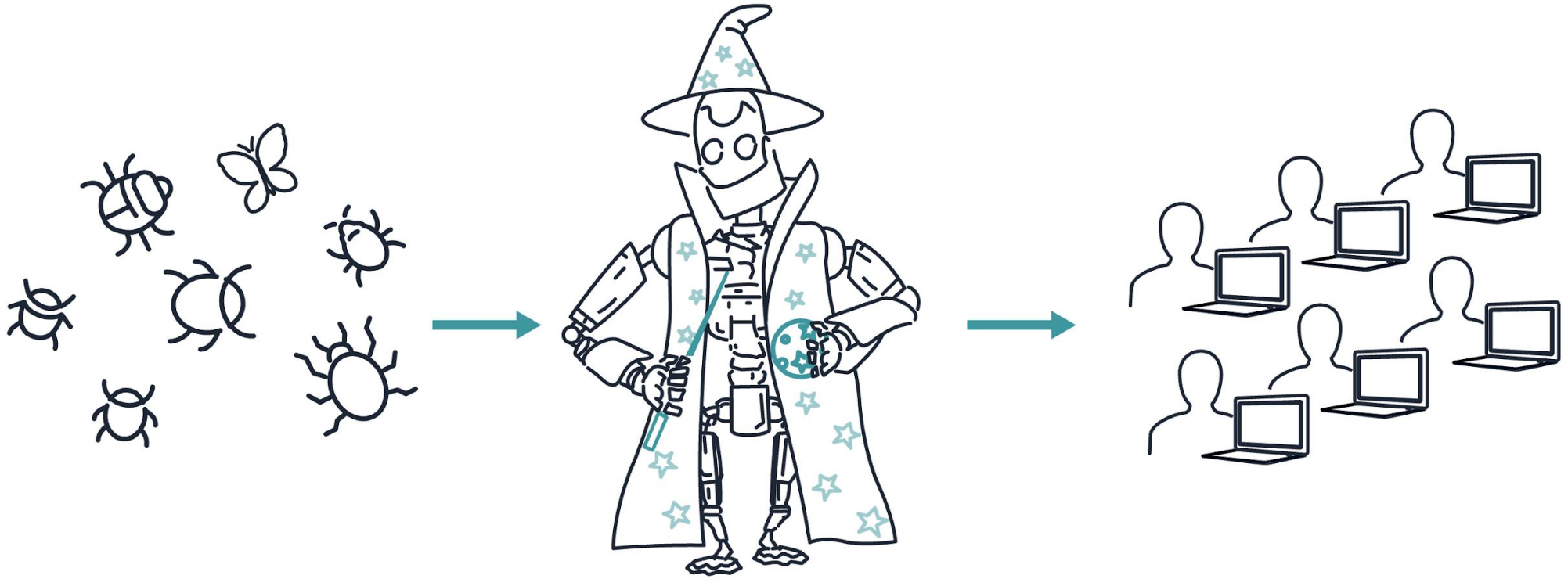
Anna Gromova
20th June 2018

Defect management

The National Institute of Standards and Technology estimated that software defects cost the U.S. economy in the area of $60 billion a year. The National Institute study also found that identifying and correcting these defects earlier could result in upwards of

$22 billion a year in savings.

Nostradamus



Increasing the Quality of Defect Reports

Nostradamus generates automatic recommendations:

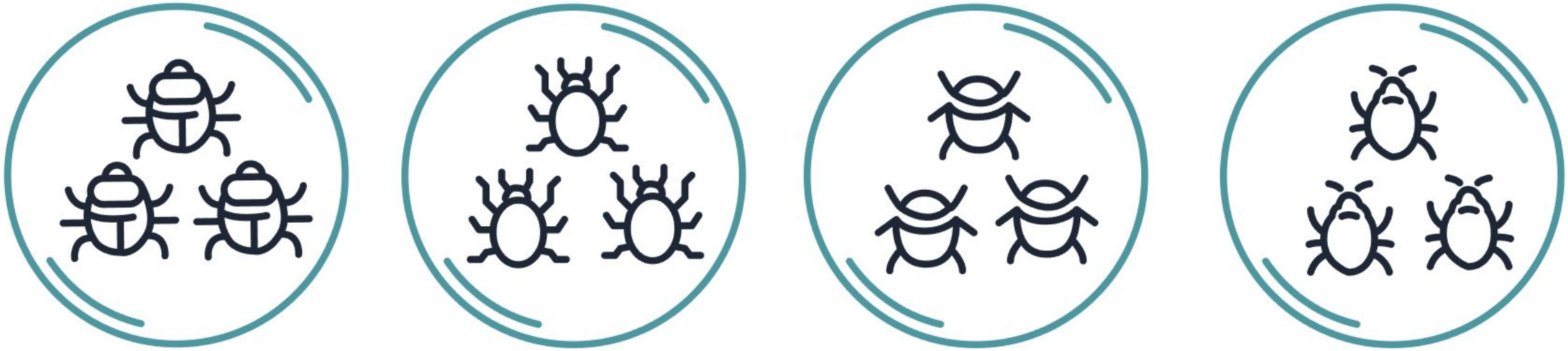
- probability of a certain priority
- probability of the area of testing
- probability of a bug being fixed (including time to resolve) or rejected

Correcting the Release Policy

Nostradamus predicts the testing metrics:

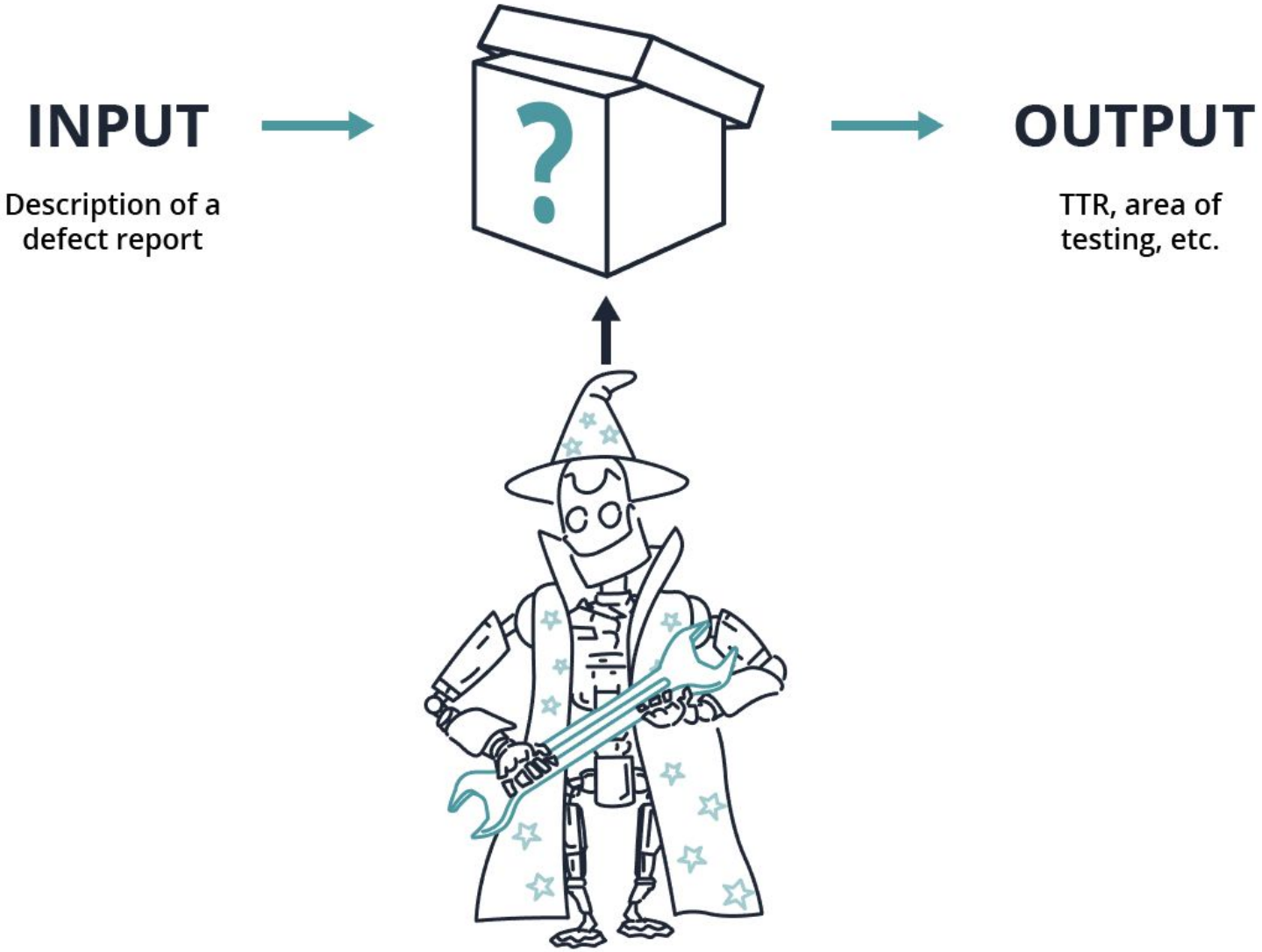
- time to fix / time to resolve (TTR)
- which defects get fixed
- which defects get rejected

Discovering Dependencies



Clustering → **Understanding the nature of defects**

Machine Learning Capabilities



Significant words

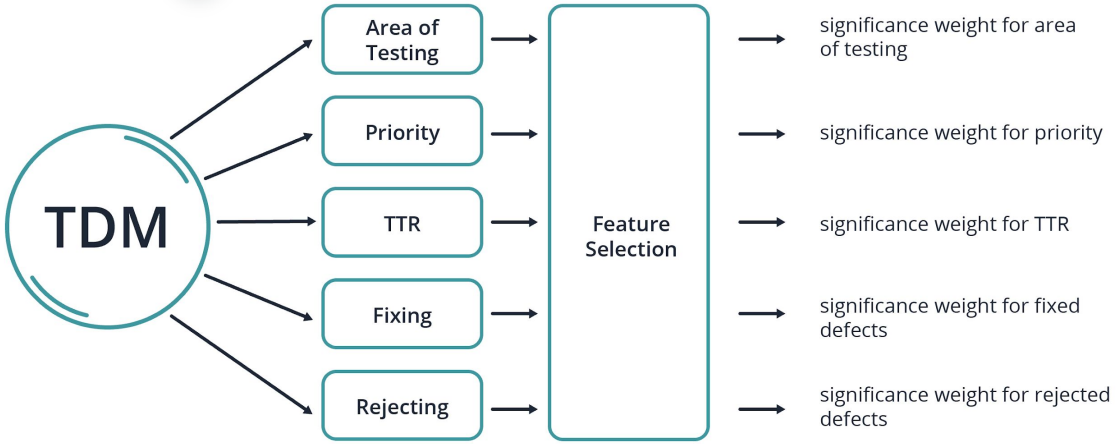
Term-Document Matrix (TDM)

TEXT →

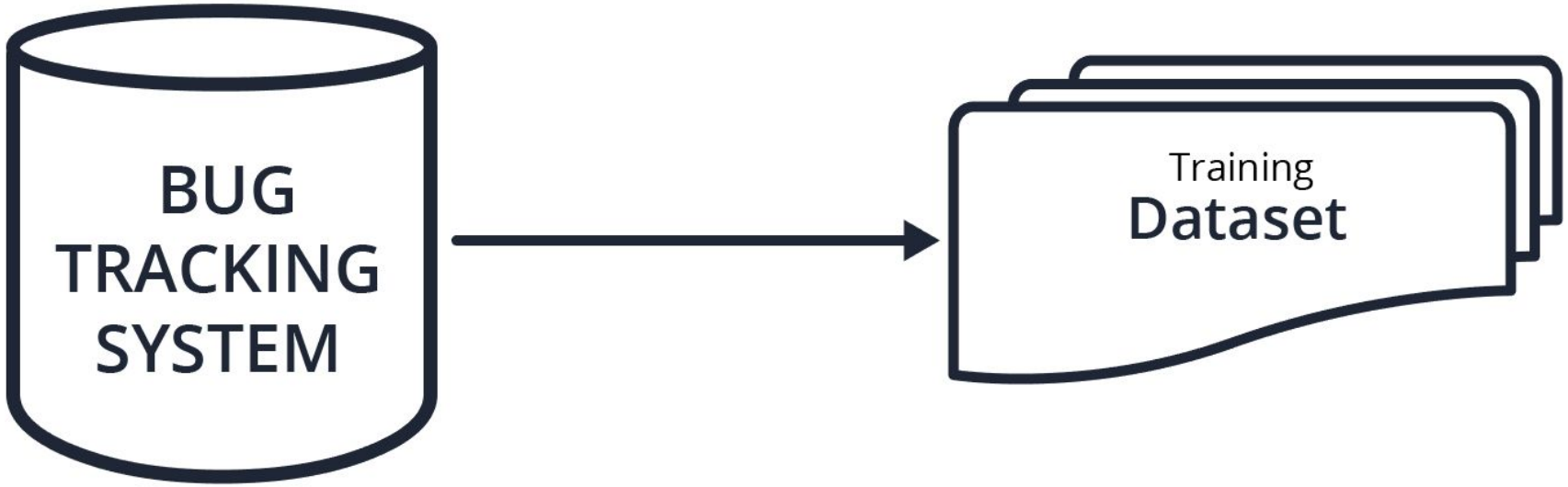
	Term 1	....	Term n
Defect report 1	weight 11		weight 1n
...			
Defect report m	weight m1	...	weight mn

→ DIGITS

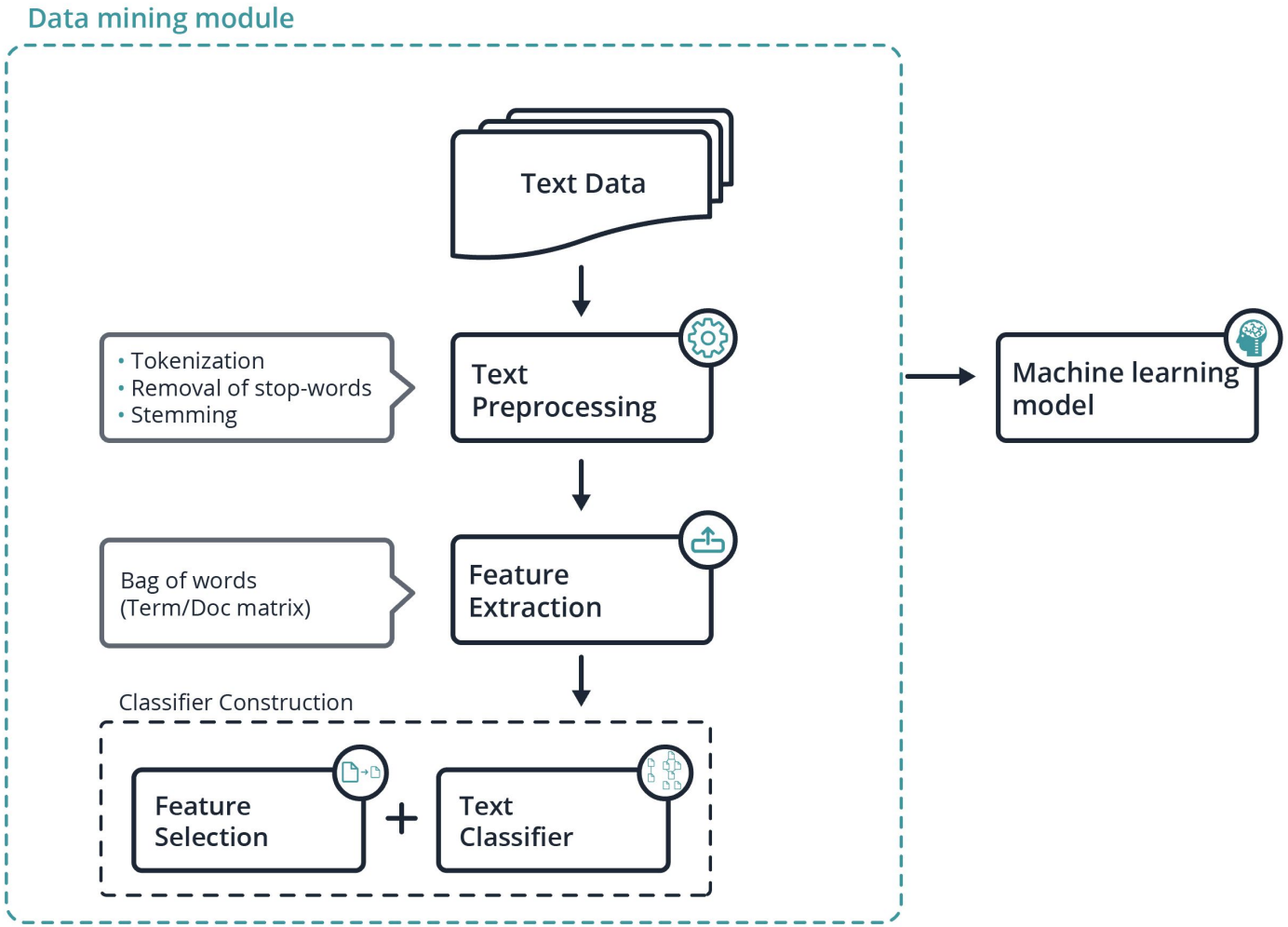
frequency weight



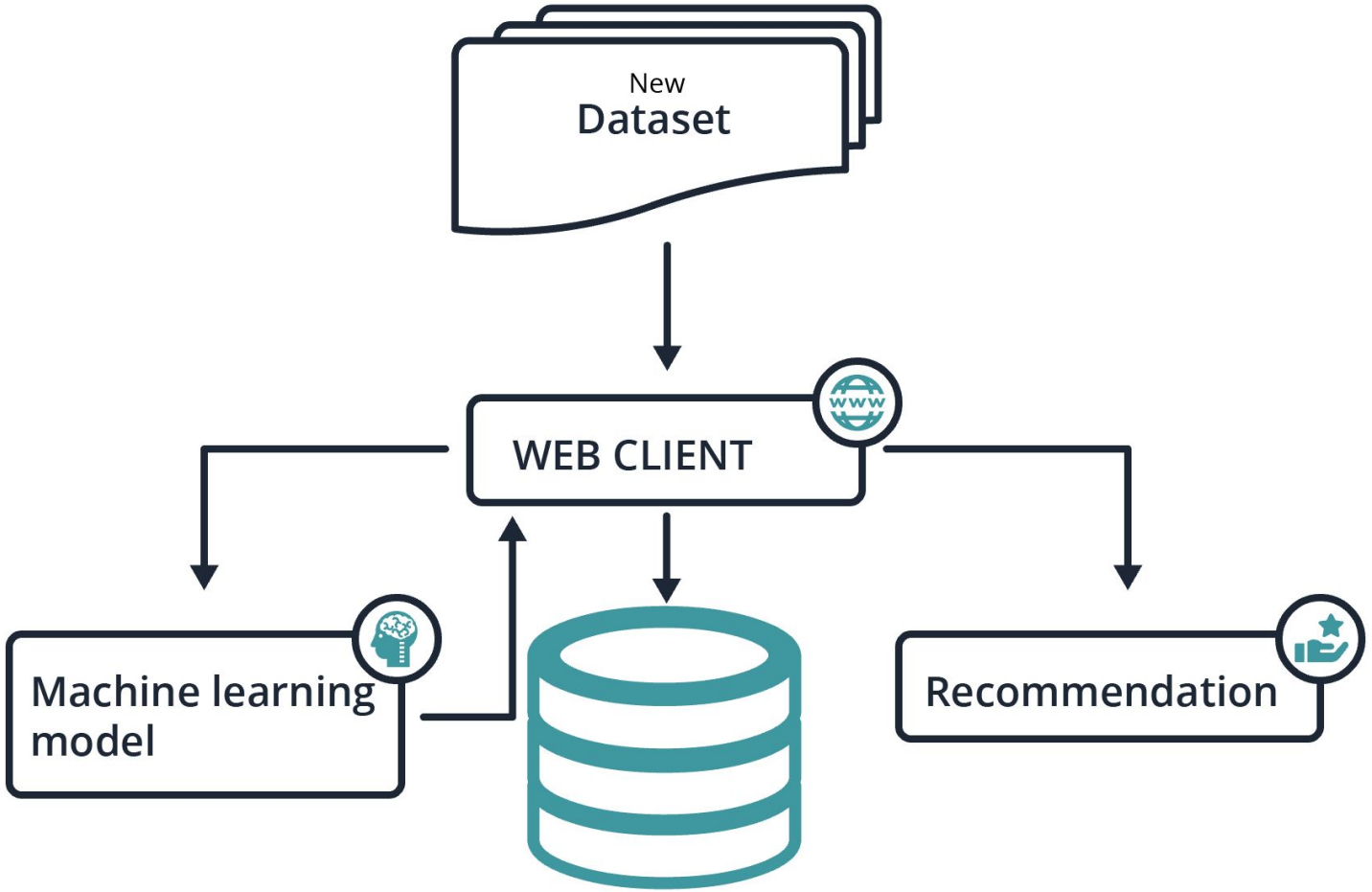
Architecture: Data Source



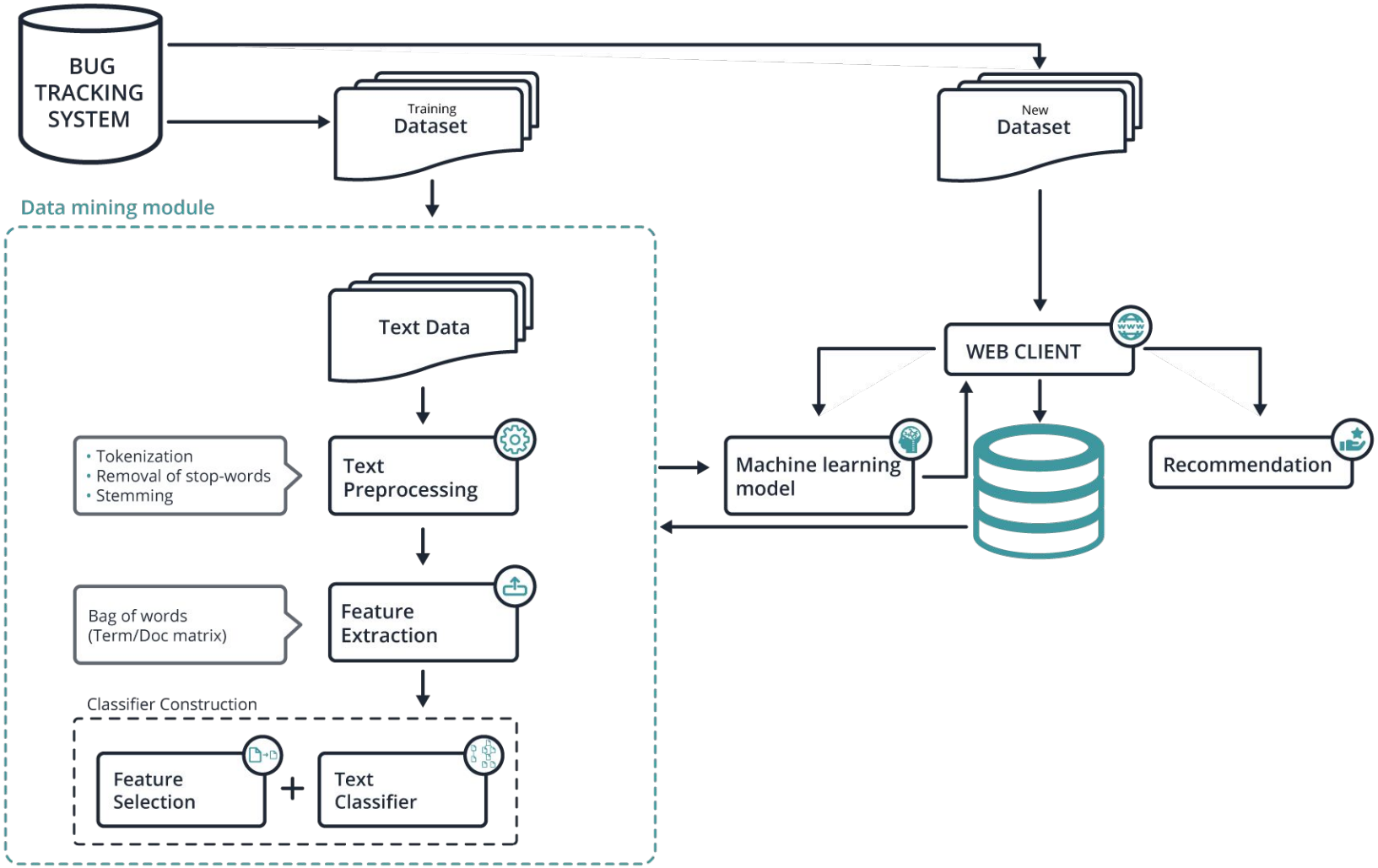
Architecture: Data Mining Module



Architecture: new data



Architecture



Technology stack

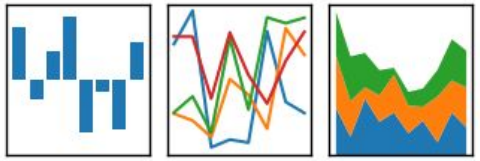
EXTENT CONFERENCE

python

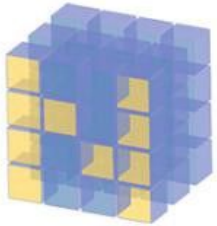
Natural Language Analyses with NLTK

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



scikit learn



NumPy

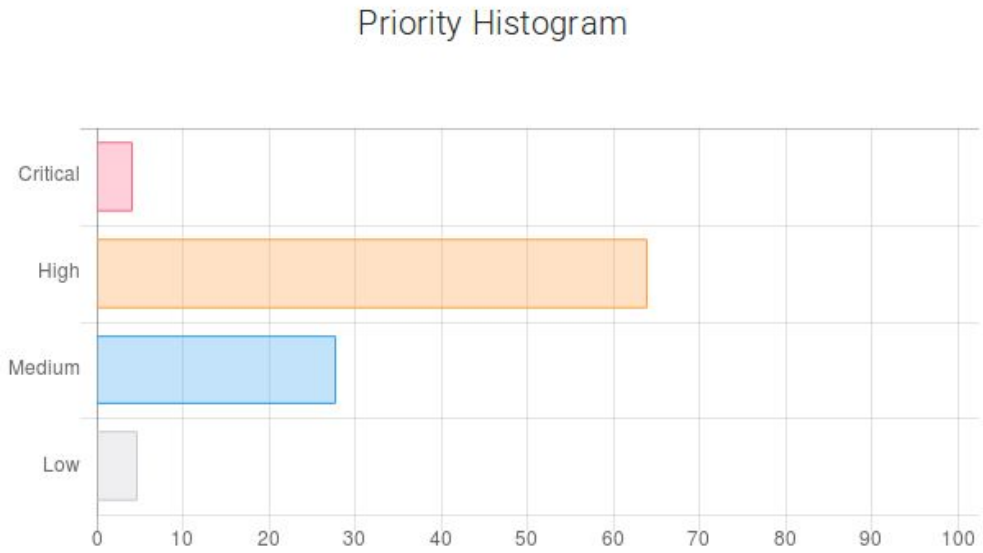
Flask

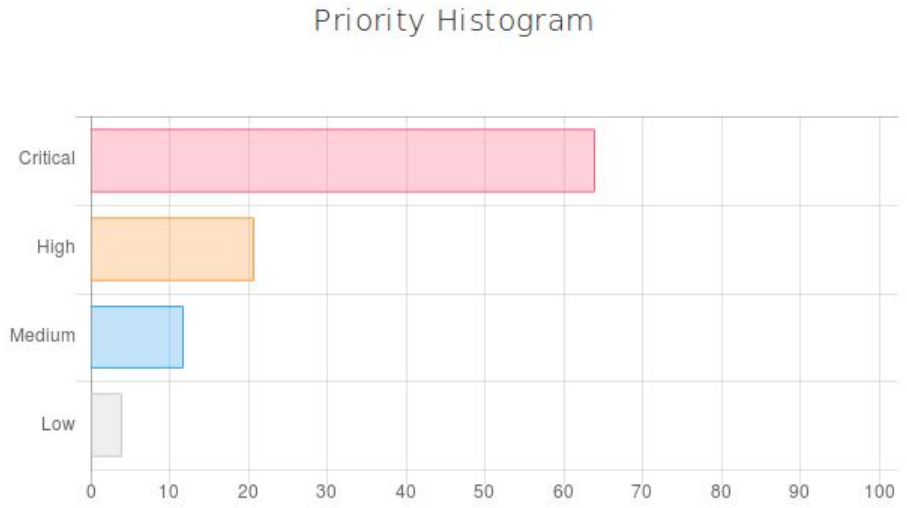
web development, one drop at a time

SQLite

Example1: Priority Changes

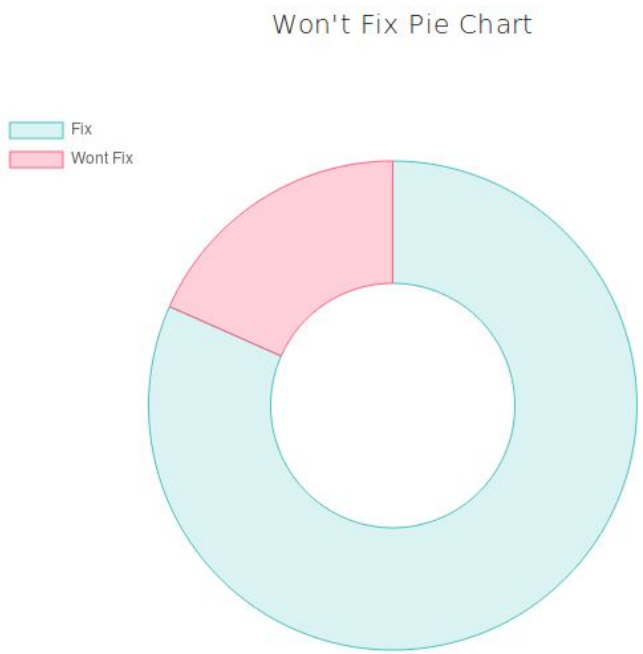
BEFORE	AFTER
Extremely high memory consumption was observed in Prod system with release 1.2.3.4.5.6 installed during ITR and ITCH gateways crashed.	*Extremely high memory consumption was observed in Prod system with release 1.2.3.4.5.6 installed during ITR. All ITCH gateways consumed up to *** GB of RAM. They crashed, except for three pairs of ITCH gateways. Probably this issue can be related to issue #1234567. Backend logs, data files, corefiles, DB dumps are attached.*





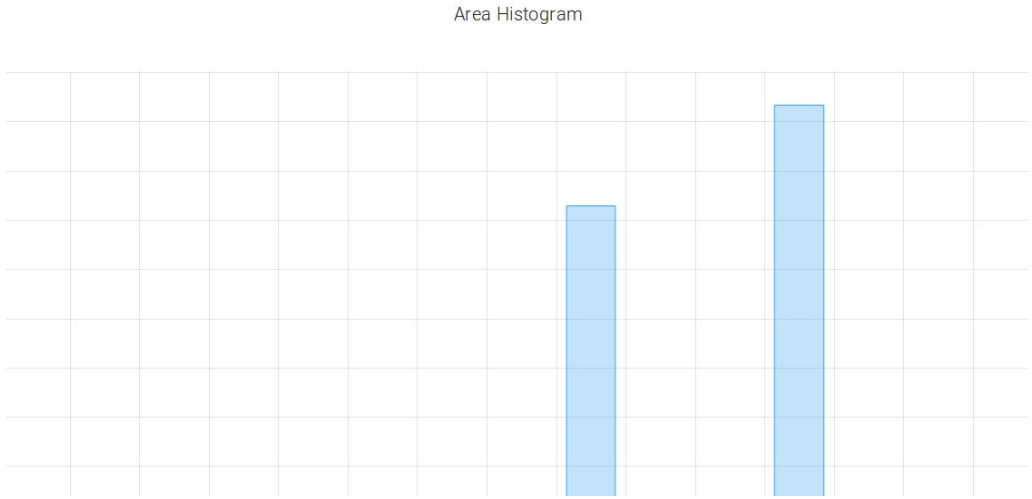
Example1: Won’t Fix Resolution Changes

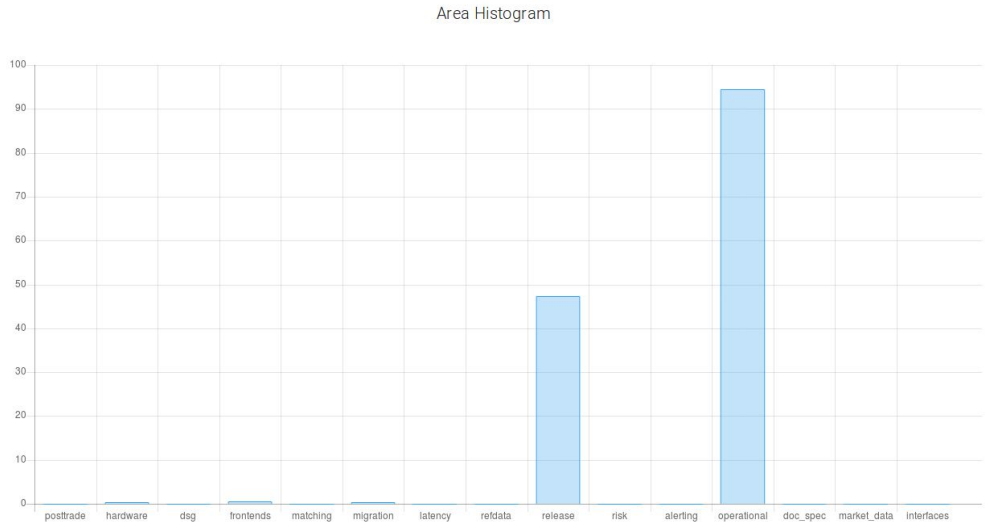
BEFORE	AFTER
Extremely high memory consumption was observed in Prod system with release 1.2.3.4.5.6 installed during ITR and ITCH gateways crashed.	*Extremely high memory consumption was observed in Prod system with release 1.2.3.4.5.6 installed during ITR. All ITCH gateways consumed up to *** GB of RAM. They crashed, except for three pairs of ITCH gateways. Probably this issue can be related to issue #1234567. Backend logs, data files, corefiles, DB dumps are attached.*



Example1: Area of Testing Changes

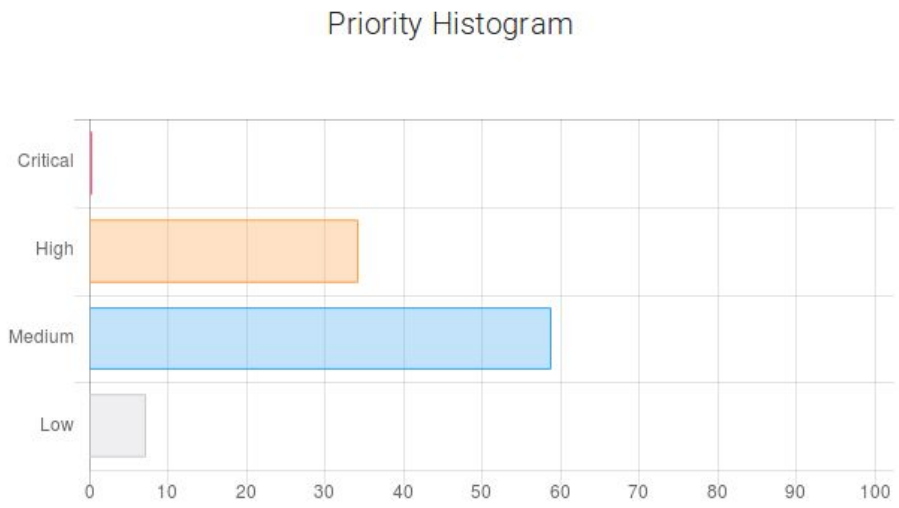
BEFORE	AFTER
Extremely high memory consumption was observed in Prod system with release 1.2.3.4.5.6 installed during ITR and ITCH gateways crashed.	*Extremely high memory consumption was observed in Prod system with release 1.2.3.4.5.6 installed during ITR. All ITCH gateways consumed up to *** GB of RAM. They crashed, except for three pairs of ITCH gateways. Probably this issue can be related to issue #1234567. Backend logs, data files, corefiles, DB dumps are attached.*

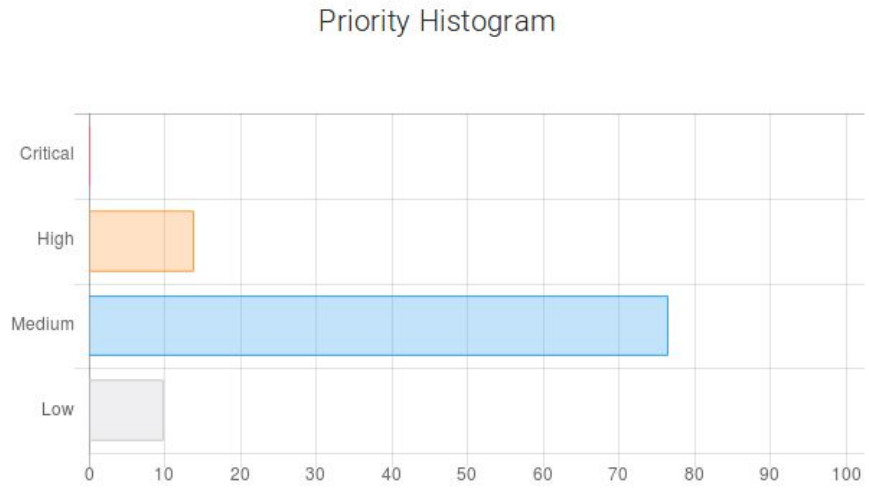




Example 2: Priority Changes

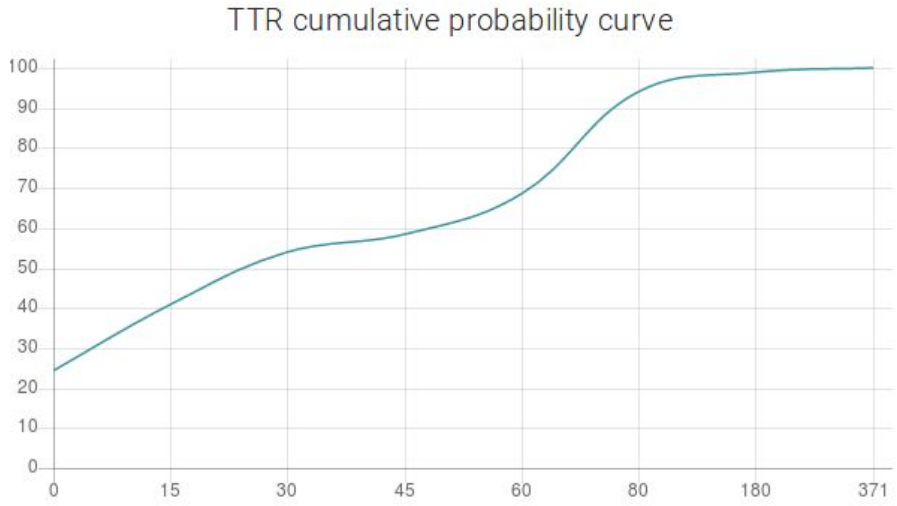
BEFORE	AFTER
The FIRM_ID column has been added to table in database in release 1.2.3.4.5. This column has value 'null' for all venues. In FrontEnds this field is displayed as not active for filling 'Participant ID'. Could you please confirm that this is expected change for database and clarify usage of this field ?	*The FIRM_ID column has been added to table in database in release 1.2.3.4.5. This column has value 'null' for all venues. In FrontEnds this field is displayed as not active for filling 'Participant ID'. But there is no description of this field in Volume 1 and 2. Could you please update specification regarding this field ?*

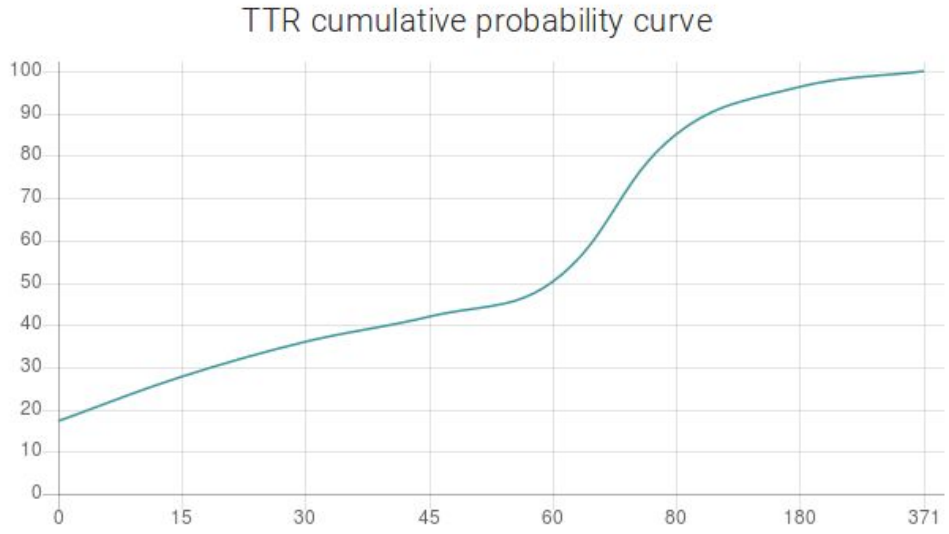




Example 2: TTR Changes

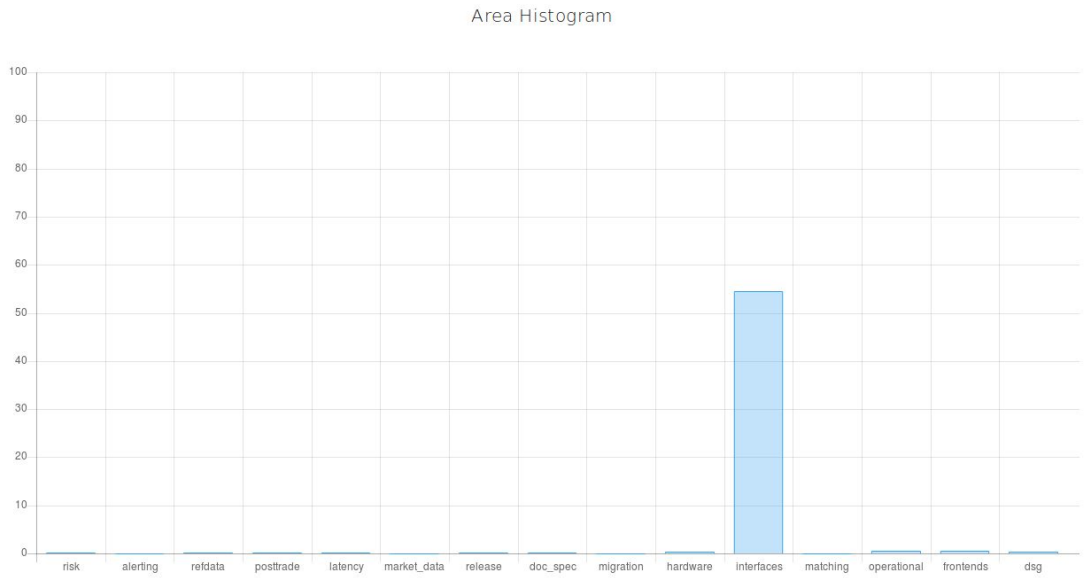
BEFORE	AFTER
The FIRM_ID column has been added to table in database in release 1.2.3.4.5. This column has value 'null' for all venues. In FrontEnds this field is displayed as not active for filling 'Participant ID'. Could you please confirm that this is expected change for database and clarify usage of this field ?	*The FIRM_ID column has been added to table in database in release 1.2.3.4.5. This column has value 'null' for all venues. In FrontEnds this field is displayed as not active for filling 'Participant ID'. But there is no description of this field in Volume 1 and 2. Could you please update specification regarding this field ?*

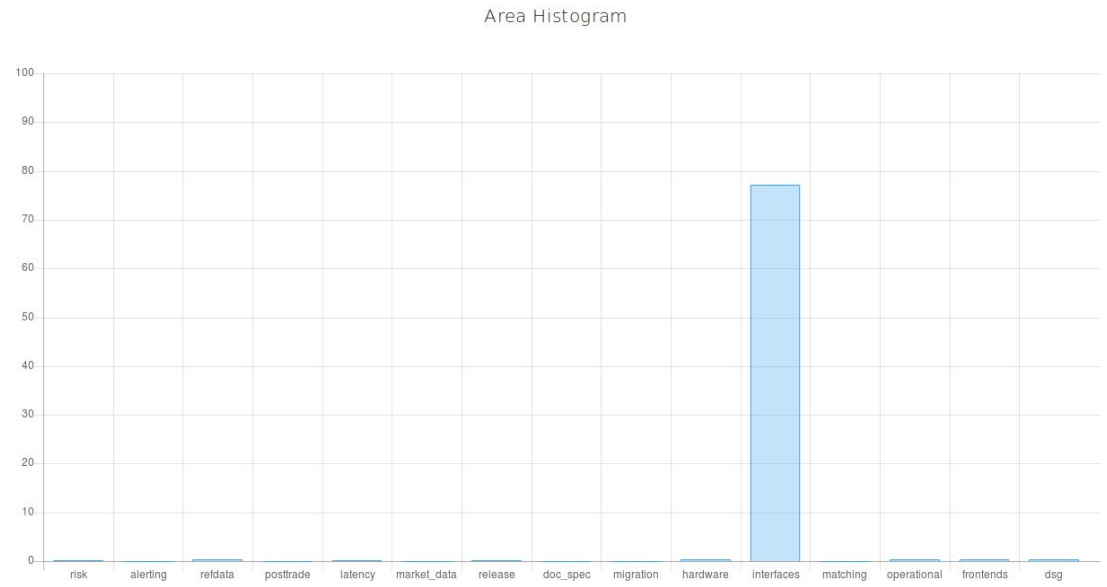




Example 3: Area of Testing Changes

BEFORE	AFTER
According to FIX Trading Gateway specification, "DeliverToCompID" is a header tag but we can see that in Execution Reports this tag is placed in the body of the message and not in its header. The issue can be observed in 1.2.2 and not reproducible in 1.2.1.	*According to FIX Trading Gateway specification "DeliverToCompID" is a header tag but we can see that in Execution Reports this tag is placed in the body of the message and not in its header. The issue can be observed in 1.2.2 and not reproducible in 1.2.1. Please refer to the attached logs and pcap files.*





Future Work

Using dynamic data to discover more complex dependencies

EXTENT CONFERENCE

Thank you!