

# Using ML to auto-classify database incidents

Henry Zhu

Special thanks to my mentor Heheng Li and my manager David Chen

# Project Overview

Use ML to classify descriptions incidents/crashes into similar groups.

Technologies Used: Python w/ Scikit-Learn Library

Procedure:

1. Preprocess data: lowercasing, lemmatization, normalization, removing stop words, HTML tags, number, and punctuation, shuffling the dataset
2. Convert descriptions to vectors using Word2Vec
3. Cluster vectors into individual groups using K-Means Algorithm

# Input

Test IcM auto flow
Test IcM auto flow
ITA-EP Production SQL Job failure   ProcessECMROCSecurityGroups
OMS Alert   Free Space Below 10%
OMS Alert   Free Space Below 10%
OMS Alert   SQL Server - SQL Server - Database BACKUP failed
Pursuit SG Addition
OMS Alert   SQL AlwaysOn/FailoverClustering Availability failure

# Output

ITA-EP Production SQL Job failure | ProcessECMROCSecurityGroups

ITA-EP Production SQL Job failure | WatermarkREsultDel

ITAuthorize PROD SQL Job failure | \_SQL\_BackupAll

ITAuthorize EP: SQL Alerts | ITAuthorizeEP PROD Reporting Replication LogReader Agent Failure

OMS Alert | Free Space Below 10%

OMS Alert | Free Space Below 10%

OMS ALert | Free Space Below 10%

FW: INC19585161 OMS Alert | Free Space Below 10%