DA592 - Term Project Proposal
May 28,2017

Overview

01 Summary

02 Motivation

03 Project Description

04 Methodology

05 Deliverables

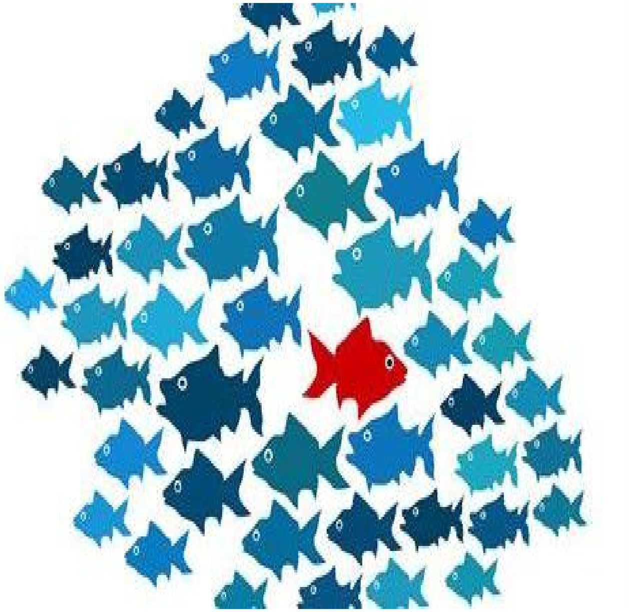
06 Milestones

01 Summary

Title: Log Anomaly Detection on Streaming Data

Start Date: Jun 02, 2017

Finish Date: Sep 22, 2017



Project Team:

- Emre Yazıcı
- Özkan Öncü
- Uğur Üker

02 Motivation

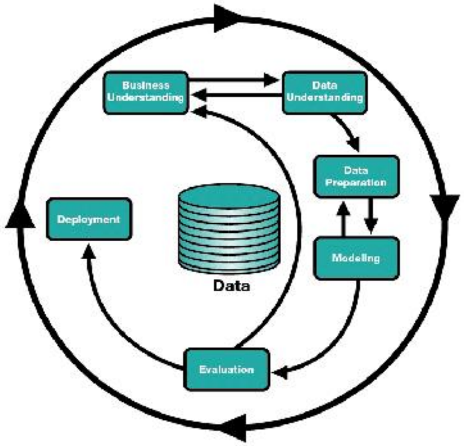
In a large environment the sheer volume of log files and events present issues and anomalies; many gigabytes of logs could be generated each day from one system alone, and thus manually reviewing log files is not only an inefficient method of review, it is also impractical. In addition, in a critical environment, some issues and anomalies must be detected whenever it happens to take immediate actions. Therefore, we would like to focus on this problem and develop a log file anomaly detector which work on streaming data. We anticipate that the application we will develop can be an alternative solution in the market against the problem.

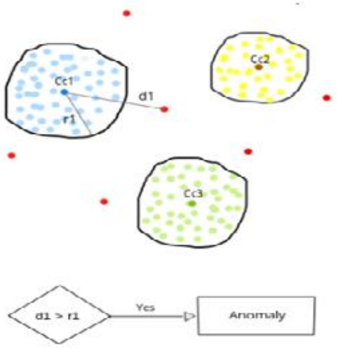
03 Project Description

It is well-known that most of corporations have very complex environment with many IT systems and IT services. Business continuity depends on keeping those systems and services up all the time. It also requires secure, reliable, and fast systems and services. Each system and service has its own pattern of activity. All those activities generate large log files every day. Monitoring and tracking the log messages are vital to provide secure, reliable and fast environment. The aim of this project is to develop an UX based application which will cluster log messages as soon as they occurs and to detect anomalies in streaming log messages immediately to warn administrator of systems and services.

04 Methodology

We follow CRISP – DM steps along project lifecycle and intend to use adaptive DBScan algorithm for clustering and to detect anomalies in streaming data. We will collect and analyze data by using Python with Apache Spark & Apache Kafka frameworks. We will probably store cluster information in MongoDB . In addition, we will apply UX Design standarts to have a easy-to-use product.





05 Deliverables

✓	Project Plan & Deliverable List & Milestones	May 25
✓	Scope Document & Literature Review Results	Jun 8
✓	Analysis & Design Documents	Jun 20
✓	Initial Prototype	Jun 30
✓	Test Results	Aug 25
✓	User Guides, Final Report & Presentation	Sep 15

06 Milestones

	Start Date - Finish Date	Status
□ Problem Definition & Kick-off Meeting	May 25 – May 25	Completed
□ Requirement Definition & Literature Review	May 28 – Jun 08	In Progress
□ Requirement Analysis	Jun 09 – Jun 20	Not Started
□ Algorithm Customization	Jun 15 – Jun 20	Not Started
□ Creating a Prototype	Jun 20 – Jun 30	Not Started
□ Development	Jul 01 – Aug 10	Not Started
□ Testing	Aug 10 – Aug 25	Not Started
□ Deployment	Aug 25 – Aug 31	Not Started
□ User Guides,Presentation & Report Completion	Aug 31 – Sep 15	Not Started
□ Project Closure	Sep 15 – Sep 22	Not Started