UNIVERSITY OF SOUTHERN DENMARK

SOFTWARE ENGINEERING 6. SEMESTER

Datamining and its use

Author:
Lasse Bjørn HANSEN
Simon FLENSTED

Supervisor:
Jan Corfixen Sørensen SØRENSEN

A report submitted in fulfillment of the requirements
of Software Engineering 6. semester

at

University of Southern Denmark
TEK

February 7, 2017

"Some quote"

- Gruppe 3

Contents

Chapter 1

Introduction

1.1 Motivation

The amount of data being processed around the Internet and within big systems is continuously increasing. This data should be structured and modelled in a way that makes it easily accessible and easy to work with. By handling these large amounts of data the right way, they can prove to be very useful, not only to the company who possess them, but also to the end users of a product. To achieve this, the art of data mining is very useful. The company Struct A/S have provided a typical software engineering task where data mining will create the foundation. This report will address theoretical aspects about data mining, how it is used in practice and how the final results of the processed data can be put to use.

1.1.1 Data mining

Data mining has become a big part of modern software engineering. Lots of companies

Chapter 2

Problem statement

2.1 Problem description

The initial problem/challenge is given to us by the company Struct A/S and is described as follows:

When launching sites, whether it being regular websites or web shops, a lot of user activity is logged. We therefore have a large amount of data associated with each of our sites but do not currently use it.
In the future we would like to be able to use logged data to generate an insight into the user activity on our site and actively use this data to create a personalized experience for the users.
The personalized experience can be in the form of product recommendations based on the users activity on the site.

2.2 Problem statement

The data we have been given is in a de-normalized format and the problem therefore comes with two challenges - normalizing the data and utilizing the data to create a personalized experience for the users.
This leads to the following research questions:

- How can you effectively normalize large amounts of data?
- How can you optimally store and access data in a scalable way?
- How can you effectively analyze large amount of data and draw conclusions from it?