

**Cluster and Cloud Computing COMP90024**

**Semester 1 / 2018**

**Assignment 2 - Australian Social Media Analytics**

City??

Team 42:

Zijian Wang <Student ID>

Ivan Ken Weng Chee <736901>

Thuy Ngoc Ha <963370>

Lan Zhou <Student ID>

Duer Wang <824325>

Table of Contents

A description of the system functionalities, the scenarios supported and why, together with

graphical results, e.g. pie-charts/graphs of Tweet analysis and snapshots of the web

apps/maps displaying certain Tweet scenarios;

● A simple user guide for testing (including system deployment and end user invocation/usage

of the systems);

● System design and architecture and how/why this was chosen;

● A discussion on the pros and cons of the NeCTAR Research Cloud and tools and processes for image creation and deployment;

● Teams should also produce a video of their system that is uploaded to YouTube (these videos

can last longer than the NeCTAR deployments unfortunately!);

● Reports should also include a link to the source code (github or bitbucket).

It is important to put your collective team details (team, city, names, surnames, student ids) in:

● the head page of the report;

● as a header in each of the files of the software project.

1. Introduction

Ref: <https://www.business.qld.gov.au/running-business/marketing-sales/marketing-promotion/online-marketing/twitter/who>

<https://www.socialmedianews.com.au/social-media-statistics-australia-february-2018/>

According to Vivid Social, a specialist advisory service for business strategies with social media, 19% of Australian internet users are using Twitter, accounted for 2.9 million monthly active user profiles (Feb 2018). The data suggests business opportunities to utilise consumer social media habits to promote sales that suit customers’ needs. This not only benefits large corporations but also small local businesses.

Inspired by the idea, our study focuses on supporting market research for food suppliers by providing insights to predict the foods that people are likely to consume using location, time, sentiments, number of friends, and whether they are homeless. By using our demographics, companies may gain quality understandings of customer behaviours and preferences to adjust existing services as well as develop new initiative products.

Our statistics are based on tweets collected from Twitter users living in Australian cities using harvester tasks running on four instances on NecTAR Research Cloud. Next leveraging CouchDB, MapReduce, Google Cloud API and Python libraries, under classification and sentiment analysis, the raw data is analyzed together with the data from AURIN focusing on our topic of foods according to selective analytic scenarios. Finally, classified data will then be visualised on our team website connected to our server by the web service module.

In this report, we will discuss the findings for three following scenarios:

* Scenario 1:
* Scenario 2:
* Scenario 3:

1. to predict the food people eat from other information

2. to predict the future trend of homeless with other information

By studying the scenarios, it is observed that <summary of findings>

1. User Guide for Testing

2.1. System Deployment

2.2. Invocation and Usage

1. System Design and Architecture

3.1. Design

Flow chart

3.2. Functionalities

1. Advantages and Disadvantages of of the NeCTAR Research Cloud

4.1. Image Creation

4.2. Deployment

4.3. Error Handling (Issues and Challenges)

1. Analytic Scenarios and Findings

5.1. Scenario Selection

Overall introduction to topic: Why and how

5.2. Scenario 1

Screenshots and graphs/charts

5.3. Scenario 2

5.4. Scenario 3

5.5. Summary of Findings

1. Conclusion and URLs

Links of Youtube video and source codes (GitHub?)

1. References