Early Rumour Detection in Social Media using Machine Learning

PSG 16 – Dr Zhiwei Lin – BEng Software Engineering

Thomas Franklin B00607399

April 2018

# Abstract

In recent times there has a been a rise in what is deemed as “fake news” when it comes to social media sites such as Twitter. Fake news can be described as the spreading of misinformation to influence the readers thoughts and is becoming a much larger issue in recent times with the rise of social media and particular “influencers”.

As social media tends to be an area where users can express their own belief, large influencers can use it as a playground to misinform their followers and hence start the process of spreading fake news – since content on social media can reach hundreds of thousands of users in minutes, which means current methods of combating fake news is often to slow to stop the spread before it starts, it was time to look in to solutions which can aid in detecting fake news a lot faster and prevent the spread before it is too late.

The aim of this project was to use machine learning to classify all public Tweets from Twitter as a rumour or non-rumour where the results could then be retrieved through a rich internet application for visualisation. From the research conducted there were very few existing projects that have tackled rumour detection in social media, and from the information gathered this will be the first project which uses natural language processing and machine learning for the early detection of a rumour in social media.

The final product produced is a service which continually streams Tweets from Twitter based on keywords, puts them through a processing process which classifies them as either as rumour or non-rumour which is then stored in a SQL database for later retrieval on an interactive web service which will show a breakdown of the results in graphs, charts etc.

In conclusion, the use of natural language processing and machine learning is only the start of tackling fake news, on its own it reduces the amount of results which will still need accessed, as the current methods where individuals who cross-verified results having to go through everything – by reducing the results earlier in the process it can speed up the cross-checking steps that will need to be done.

# Acknowledgements

This project would not have been possible without the support of my mentor Dr. Zhiwei Lin for his guidance and support throughout. Another thanks goes towards the Ulster University teaching staff of whom I had the pleasure to learn from over the last 4 years, and finally I would like to thank those I had the pleasure of working with while on placement – most importantly Aaron Long who was a mentor to me when starting out in placement and with his continued support I have become a better developer.

**Table of Contents**

[Abstract 1](#_Toc511126737)

[Acknowledgements 1](#_Toc511126738)

[1. Introduction 3](#_Toc511126739)

[1.1 Problem Elucidation and Statement 3](#_Toc511126740)

[1.2 Project Aim 3](#_Toc511126741)

[1.3 Project Objectives 3](#_Toc511126742)

[1.4 Selected Software Lifecycle Methodology 4](#_Toc511126743)

[2. Requirement Control Document 4](#_Toc511126744)

[2.1 Requirements Gathering Methodology 4](#_Toc511126745)

[2.2 Final Requirements 5](#_Toc511126746)

[2.3 Requirements Evolution 6](#_Toc511126747)

[3. System Overview 6](#_Toc511126748)

[3.1 System Design 7](#_Toc511126749)

[3.2 Interface Design 8](#_Toc511126750)

[3.2.1 HCI Considerations 8](#_Toc511126751)

[3.2.2 Wireframes 9](#_Toc511126752)

[3.3 Data Support Design 10](#_Toc511126753)

[3.3.1 Consideration for Security and Data Validation 11](#_Toc511126754)

[3.3.2 Database Design and Related Artefacts 12](#_Toc511126755)

[3.4 User Interaction Design 14](#_Toc511126756)

[3.4.1 Information Flow Diagram 14](#_Toc511126757)

[3.4.2 Use Case Diagrams 15](#_Toc511126758)

[3.4.3 Activity Flow Diagram 16](#_Toc511126759)

[3.5 Other Design Artefacts 17](#_Toc511126760)

[Bibliography: 18](#_Toc511126761)

# Introduction

This report will outline the solution developed to tackle the problem identified in the report, along with describing the development process of the proposed solution along with the management and elicitation of the product and user requirements which will be documented in the report, the report will also highlight the design and testing which had been completed to ensure that the project is a success.

## Problem Elucidation and Statement

Fake news[[1]](#footnote-1) is a term which has risen in popularity within the last few years, especially when it comes to social media sites such as Twitter and Facebook (Meyer, R, 2018; Science | AAAS, 2018), although in recent months there has been responses from Facebook on their plans to tackle fake news (The Guardian, 2018; Bloomberg.com, 2018), Twitter on the other hand has been less vocal in their attempts and have only suspended accounts and emailed those who may have been in contact (Ecommerce Week. 2018).

Evidence has come to light in recent months about how the 2016 US Presidential Elections were influenced by what is termed fake news, and how the Russians pioneered and used it to influence the election (Vox, 2018; Conger K, 2018), had there been detection measures in place the spread of the misinformation could have been stopped early enough in order that it had no impact on the results.

Should nothing be done to try and reduce, or completely stop the spread of misinformation then it could continue to sabotage future important events, and as out of the two main social media sites, only Facebook has begun to address the issue it leaves an area open in Twitter to implement a solution to help in the early detection of rumours before they have a chance to spread and sabotage important events.

Should the project be capable of aiding in the early detection of social media the project code should be reusable for different functions that too require assistance in the detection of rumours, which gives it large commercial value as any company could reuse the system with their own filters to find out users opinion on themselves and if users are spreading misinformation about their products which could be brand damaging where they could intervene early and reduce the damage to their brand.

## Project Aim

This project aims to aid in the early detection of Tweets from Twitter early enough, in order for the appropriate action to be taken to prevent the spread of misinformation; often referred to as fake news. With the use of machine learning, this project will continuously stream Tweets from Twitter in real-time using configuration filters for areas of interest, such as Politics, Crisis’, Disasters etc., using the Twitter API which will then classify the individual Tweets as a rumour or non-rumour using the trained classifier where the results will be stored in a database which where an interactive website can get the data required to present the information using various visualisation techniques such as graphs, word clouds, etc.

## Project Objectives

Table 1 Project Objectives Table

|  |  |
| --- | --- |
| Id | Objective |
| 1 | Design and implementation of a system which can continuously process Tweets from Twitter using configurable filters |
| 2 | Implement a classifier capable of assigning a rumour or non-rumour label based on the text from a Tweet |
| 3 | Design and implementation of a website to display the information in meaningful ways using best practices |
| 4 | Design and implementation of a backend service which can retrieve the data |
| 5 | To follow a project plan to ensure completion of the project by the project deadlines |
| 6 | To evaluate different machine learning libraries which can be used for the development of the classifier in objective 2 |
| 7 | To evaluate different frameworks and libraries which can ensure a well-executed system which adheres to appropriate system design considerations |
| 8 | To implement a system which is capable of running independently without much user interaction to adhere to the real-time aspects required of the project aim |
| 9 | The design and implementation of a system which can be reused by others who wish to monitor rumours of a given topic |

## Selected Software Lifecycle Methodology

The lifecycle followed by this project was the Extreme Programming (XP) software development methodology.

# Requirement Control Document

This section will document the methods used by the project owner to gather user/product requirements, along with tables outlining the final set of functional and non-functional requirements which had been identified with the prioritisation value using the prioritisation technique known as Wiegers Relative Weighting, as well as the explanation as to how the requirements evolved during the lifecycle of the project.

## Requirements Gathering Methodology

The initial requirements of the project and for any additional requirements were gathered using the following techniques: peer reviews, scenarios and walk-throughs which were all performed with various members of the stakeholders of the project. The techniques suited the format of the project, as regular meetings were held with the stakeholders in the early stages, and through these meetings the requirements could be identified using the three techniques mentioned above and they were refined using brainstorming techniques to help shape them in to tangible requirements.

## Final Requirements

For the prioritisation of requirements, the project utilised Wiegers Relative Weighting technique (Wiegers K, 1999) which is a technique which uses cost, value and risk to prioritise the requirements. Table 2 shows the final set of functional requirements for the developed system along with their prioritisation value, while in Table 3 it shows the non-functional requirements of the system along with their prioritisation value.

The prioritisation value is calculated using Wiegers Relative Weighting formula as below:

The higher the value the higher the priority, tables 2 and 3 only show the value, they are organised by the Requirement ID rather than the prioritisation value.

Table 2 Final Functional Requirements and Prioritisation Value

|  |  |  |
| --- | --- | --- |
| ID | Description | Prioritisation Value |
| F.01 | System shall continuously process Tweets without user interaction | 0.75 |
| F.02 | Users shall be able to access the website through any browser | 0.7 |
| F.03 | Users shall be able to search for individual hashtags which they are interested in | 0.28 |
| F.04 | System shall display the top hashtags using various visualisation techniques | 0.9 |
| F.05 | System shall display the top users using various visualisation techniques | 0.9 |
| F.06 | Users shall be able to search for individual users which they are interested in | 0.39 |
| F.07 | System shall assign labels to individual Tweets either as a rumour or non-rumour | 0.9 |
| F.08 | System shall store the classification label of each individual Tweet | 0.9 |
| F.09 | System shall store location information of the Tweet | 0.23 |
| F.10 | System shall display service status information | 0.6 |
| F.11 | System shall allow user to search rumours/non-rumours near them | 0.28 |
| F.12 | System shall allow user to report a particular user | 0.23 |
| F.13 | System shall allow user to report a particular hashtag | 0.23 |
| F.14 | System shall provide help through tooltips and a help menu | 0.67 |

Table 3 Final Non-Functional Requirements and Prioritisation Value

|  |  |  |
| --- | --- | --- |
| ID | Description | Prioritisation Value |
| NF.01 | System shall not store data any longer than required | 0.67 |
| NF.02 | System shall be developed using open source frameworks and libraries to reduce risk | 0.55 |
| NF.03 | System shall be tested using appropriate testing strategies | 0.67 |
| NF.04 | System should be robust enough to restart on failure | 0.31 |
| NF.05 | The system shall be protected from SQL injection attacks | 0.45 |
| NF.06 | Systems interface shall be pleasing to the eye | 0.61 |
| NF.07 | System shall log appropriate metrics for future enhancements | 0.25 |
| NF.08 | System shall give user a response within 5 seconds of navigating to a page | 0.5 |
| NF.09 | System shall be developed using best practices | 0.67 |

## Requirements Evolution

From the initial report there have been a few additional requirements which evolved from the regular meetings with the stakeholders using the techniques identified in section 2.1. Notably requirement F.04,05 were added as they matched the capability of the system in its early prototypes, while the functional requirements relating to search (F.03,06,11) were added as they were future enhancement considerations.

As the project followed an agile lifecycle methodology there were plenty of opportunity to revisit the requirements, and add to them if required; as each iteration of the system went through various stages, one of which was release planning/sprint planning where user stories were prioritised, and as part of planning a user story the requirements had to be taken in to account as to what is being addressed – during these stages, requirements had been amended, removed or split to be less ambiguous.

# System Overview

This section will outline the system design, which includes the architecture of the system. The system was designed with performance in mind as it has to continuously stream and process Tweets in real-time while not hindering the user performance which is why the system was split in to multiple micro-services, which meant that each service only had one responsibility which allows for easy scalability of the service if required, and as the services are loosely coupled[[2]](#footnote-2) they can be maintained easier.

Along with the system design, the section will also cover the User Interface design, which will include the wireframes which were designed, and the considerations made for design best practices such as HCI[[3]](#footnote-3). As the system uses a SQL database for the storage and retrieval of classified Tweets, section 3.3 will show the design of the database, displayed using ER diagrams. Section 3.4 provides information on the system through the uses of use case diagrams and activity flow diagrams and aims to provide detail on how users will interact with the system.

## System Design

The system was designed with micro-services in mind, to ensure that the system can easily be scaled if required and also to ensure that the system can deal with the real-time aspects of the service. With all this in mind, the system was split in to two parts, one which dealt the with processing, and then one which dealt with the user interaction – this allows the system to be more robust in the fact that users interactions will not slow down/affect the processing of data, while the processing of data will not negatively impact the user. Figure 1 shows the architecture of the system where it highlights the high-level components of the system.

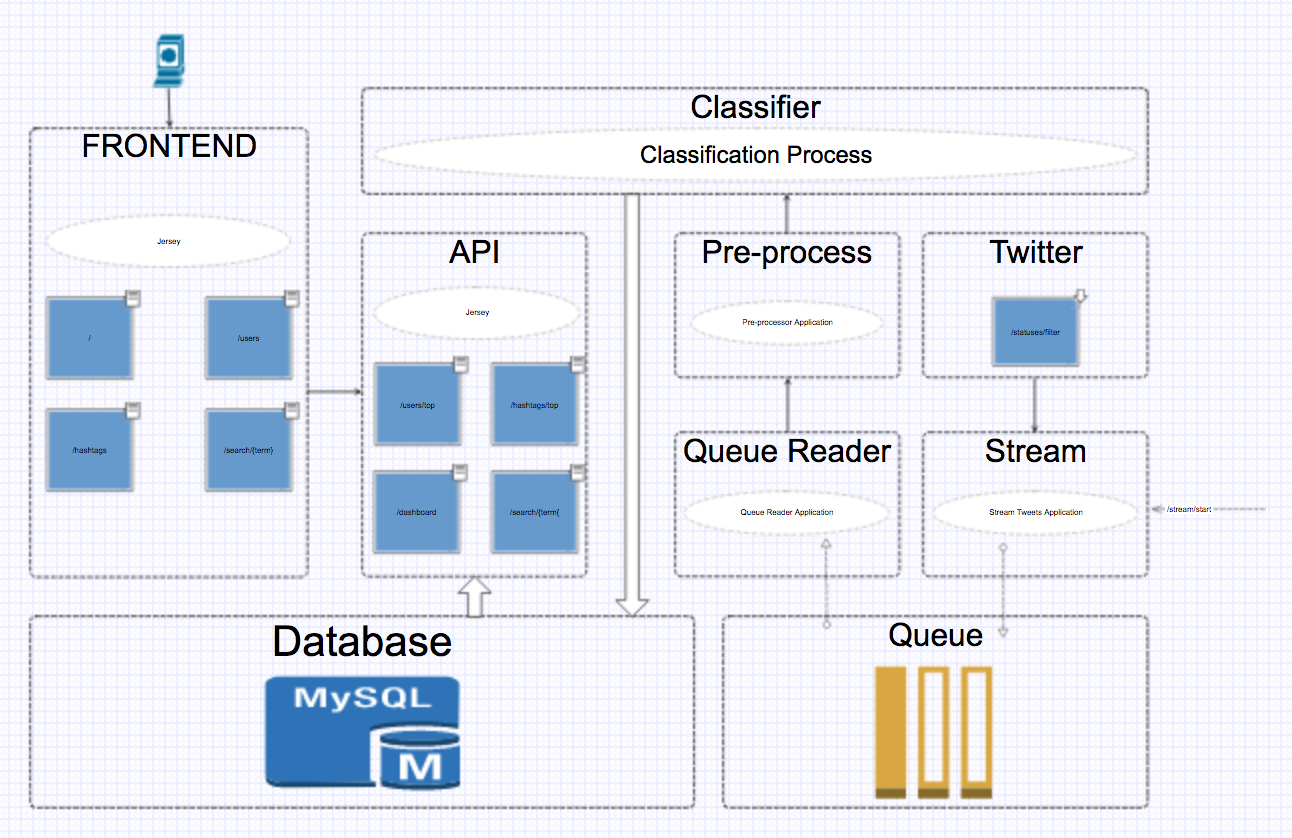


Figure 1 High-level System Architecture Diagram

From figure 1 each of the core components are listed, beginning with the one labelled *Twitter* is the Twitter API, and the *Stream* calls the API using Twitter4J[[4]](#footnote-4) which opens up a filtered stream[[5]](#footnote-5), the filter is a list of words which will be streamed in to the service, and is one of the ways in which the system can be reused as different companies can apply their own filters for topics which are relevant to them.

The *Stream* service is the control over the system, and has an endpoint exposed to start and stop the stream when required, when the stream is started it will continuously stream Tweets as per the product requirements identified in section 2.2.

From the stream, each individual streamed Tweet will be loaded on to a message queue in order to keep up with the volume of Tweets coming through the service, the service labelled *Queue Reader*, reads from the same message queue and dispatches it to the *Pre-process* service – where some pre-processing is done on the Tweet, such as lowercasing, removal of certain symbols, etc. in preparation for the classification process, using an endpoint exposed on the service labelled *Classifier* the pre-processor posts the processed Tweet item to it, this is where the Tweet body gets assigned a label as either a rumour or non-rumour based on the classification process outlined in section 3.5.1 – after the classification process, the *Classifier* saves the item to the *Database*, of which the design is outlined in section 3.3.

This is the end of what is referred to as the offline side of the service, which continuously processes in the background without user interaction, moving on to the service labelled *API* is where the user interacts with the service and is referred to as the online part of the service, and it will have various endpoints to retrieve different bits of data in a RESTful[[6]](#footnote-6) way. The service labelled *Frontend* works closely with the *API* as the *Frontend* makes use of the APIs endpoints to retrieve the required data from the *Database* to display to the user on the *Frontend*.

All the services identified in Figure 1 are all ran within their own Docker containers, as a container is a lightweight, standalone package which have all the required tools to run the service (Docker, 2018), i.e. the *Frontend* service will have Tomcat[[7]](#footnote-7) running which will expose the ports in order to connect to the web service, while *Classifier* will have Weka[[8]](#footnote-8) and Mallet[[9]](#footnote-9) running in order to perform the classification of the Tweets.

## Interface Design

The system will provide the required evidence to show that the interface design was done with best practices in mind by following the core principles of HCI and how these considerations will help to improve the usability and accessibility of the final product. Along with the evidence of following the best practices, it will also provide the wireframes that were produced as a guideline for the finalised pages of the website.

### HCI Considerations

When it comes to HCI, there are two key practices which typically come up, the 10 Usability Heuristics for User Interface Design (Jakob Nielsen, 1995) and the 8 Golden Rules of Interface Design (Ben Shneiderman's, 1986). The 10 Usability Heuristics iterate over the ‘golden rules’ (UX Courses, 2018) and as such the project has used these heuristics when considering the design of the interfaces in the project, table 4 shows the 10 heuristics with an explanation.

These 10 heuristics set the guidelines which have been followed when considering the interface of the project, and as such they aim to enhance the usability and accessibility of the service. Interface design should not end at the website, as the project relies heavily on the backend API, the interface of the API needs to be considered, which is why HCI has to be applied through the project to ensure that all aspects applies the principles set out by Jakob Nielsens 10 Usability Heuristics for User Interface Design.

Table 4 10 Usability Heuristics for Interface Design and a brief explanation

|  |  |  |
| --- | --- | --- |
| ID | Heuristic | Explanation |
| 1 | Visibility of system status | Keep users informed of system progress, i.e. use of progress bars, etc. |
| 2 | Match between system and real-world | Use language familiar to the user, i.e. use icons that represent real-world things |
| 3 | User control and freedom | Allow a “quick escape”, i.e. if user finds they’ve went too far, allow them to go back easily |
| 4 | Consistency and standards | Follow conventions and don’t use multiple words/icons to mean the same thing, i.e. use words consistently |
| 5 | Error prevention | Avoiding situations which could cause an error is better than descriptive errors, i.e. having a fault free system |
| 6 | Recognition rather than recall | Minimising users’ memory load, i.e. the user shouldn’t have to recall something from a previous page if the information is used elsewhere |
| 7 | Flexibility and efficiency of use | Allow users’ control over the system to meet their needs |
| 8 | Aesthetic and minimalist design | Only present users with information that is relevant, do not overburden them with too much |
| 9 | Help users recognise, diagnose, and recover from errors | If an error occurs, use common terminology and do not hide behind error codes |
| 10 | Help and documentation | Although it is best to develop a system which requires minimal support, if there is documentation required ensure it is easy to navigate and gives concrete steps |

### Wireframes

Wireframes are used early in a project to establish basic page structures before any visual design or content is added (Experience UX, 2018). As a wireframe is only an early visual to provide an indication of where core page content and functionality should be positioned – not all structural elements will be shown in them, and as such they are only guidelines for the final design.

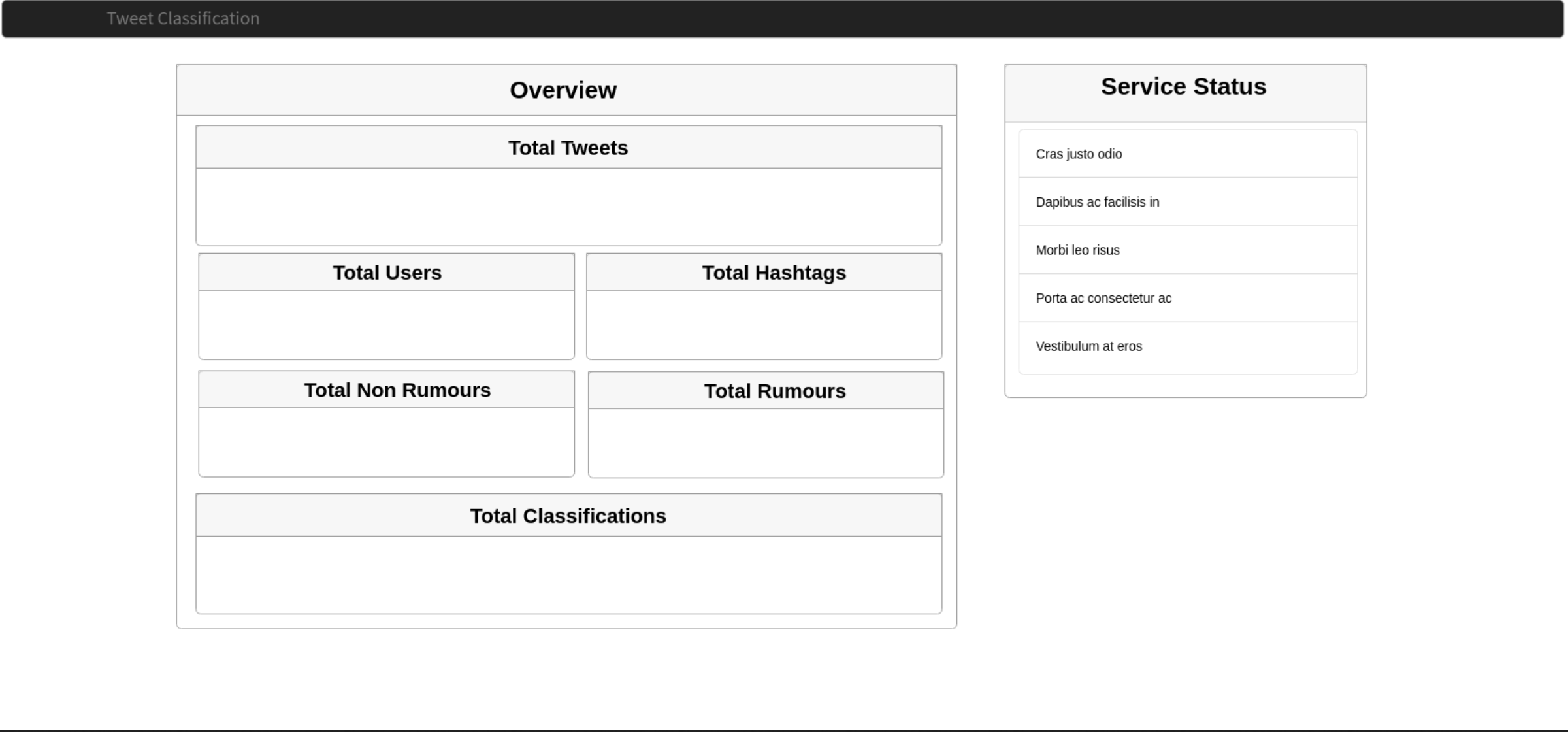


Figure 2 Wireframe of the Dashboard - the "homepage" of the website

Figure 2 is a wireframe of the dashboard of the service, which gives an overview of the data which is stored and gives the user a quick glance indication of the system status, such as which services are running – which relates back to heuristic 1 from table 4, about the guideline for providing users with information on the systems current status.

In Figure 3 it illustrates what key information could be provided when the user goes to the “hashtags” section of the service, and as such will allow the user to select different hashtags based on their rank and retrieve information specific to that hashtag – using the term “hashtag” corresponds with heuristic 4 from table 4, as the term “hashtag” is what users’ of the service will be familiar with.

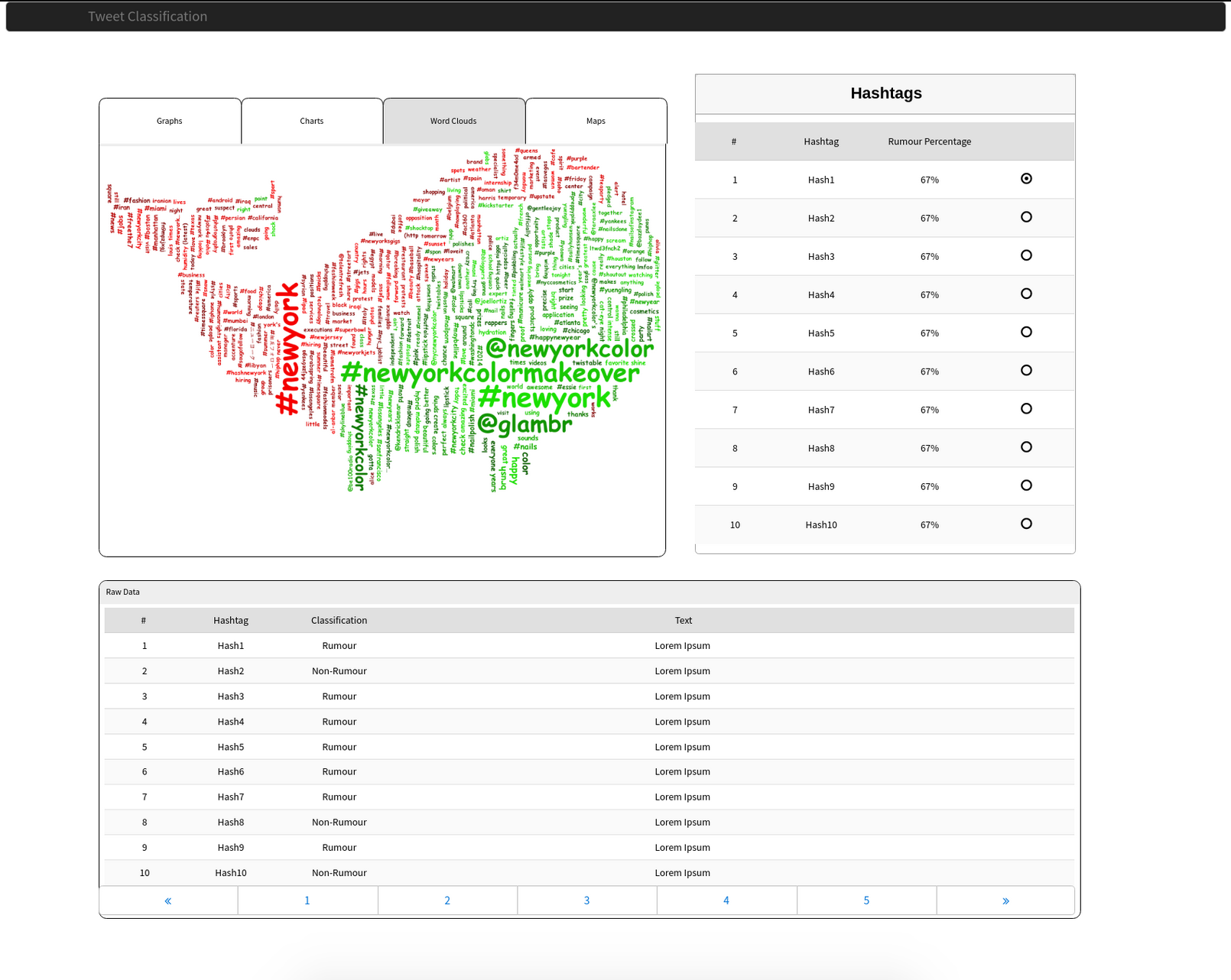


Figure 3 Top Hashtags Wireframe

As the service groups the rumours and non-rumours between users and hashtags, the wireframe in figure 3 also covers the “users” classification page where it too will show the top users rank and allow the user to see various visualisation techniques for the user data. As the classification is done on an individual Tweet, the users/hashtags are just how the data is grouped, and as such the pages for “users” and “hashtags” should be visually similar in order to adhere to heuristic 4 and 6 from table 4.

## Data Support Design

This section will aim to cover the consideration of data validation and security measures for the data of the project, along with the database design artefacts, such as the ER Diagram. As this project is very focused on the accuracy of information and the speed in which it is retrieved, getting the design of the database right is important, as a poorly design database will negatively impact the performance of a system which relies on data (Database Trends and Applications, 2018; SearchDataManagement, 2018).

### Consideration for Security and Data Validation

The most common attacks on a database fall in to the following categories; SQL Injection, Buffer Overflow vulnerabilities, Denial of Service, Privilege Escalation and Weak Authentication (Checkmarx, 2018) all of which can be avoided with good database design and few basic steps to avoid such attacks. First of all, to combat privilege escalation, creating a user and revoking all access is performed, and then only granting the required privileges to the database and tables;

REVOKE ALL PRIVILEGES ON \* FROM 'twitter'@'%';

(Revoking all privileges for everything for user ‘twitter’)

GRANT SELECT, UPDATE, DELETE, INSERT ON twitter\_classification.\* TO 'twitter';

(Only granting SELECT, UPDATE, DELETE and INSERT access to twitter\_classification database for user ‘twitter’)

By removing privileges to all other databases, the user ‘twitter’ will not be able to escalate their privileges to get more access than they are allowed. Secondly for weak authentication, using passwords of suitable length and complexity will be applied, and ensuring that the ‘root’ user has a complex password set.

As Denial of Services (DoS) attacks on a database are often cause by Buffer Overflows (Checkmarx, 2018) and are designed to halt the database, by putting rate limits on the ‘twitter’ user will help to prevent a DoS attack and buffer overflows;

ALTER USER 'twitter'@'%' WITH MAX\_QUERIES\_PER\_HOUR 5000;

(Setting the maximum queries per hour for user ‘twitter’)

Finally, for SQL Injection, this will be addressed from the API level, where no data will directly be inserted to the database as all communication with the database will be executed as a Prepared Statement.

Addressing the validity of data should be performed at all levels to ensure a smooth user experience, i.e. a user should not have to wait till they submit something to find out that there is an issue with what they entered. Although in terms of database design, data validation can be done by using a few simple techniques (Validation | Databases | ICT, 2018);

1. Type – using appropriate field types, i.e. for fields which will only contain numbers, use a numeric type, integer, big integer etc.
2. Presence – for compulsory fields, ensure that the database design acknowledges that they cannot be null or empty
3. Uniqueness – any data which should only occur once should be made as a unique field, or a combination of fields should be unique
4. Range – limit the amount of data which it can hold, i.e. if you know the text will be a maximum of 300 characters, then only allow it to store that amount
5. Format – particular inputs might need to be in a certain format

The list above is some of the techniques which can be applied in database design to ensure that data is validated correctly and in the appropriate format, which is expected by the application.

### Database Design and Related Artefacts

The project went with a SQL database as the data has strong relations between a Tweet and a Classification value which then had relationships between individual users of Twitter and could have numerous hashtags that were in the original Tweet.

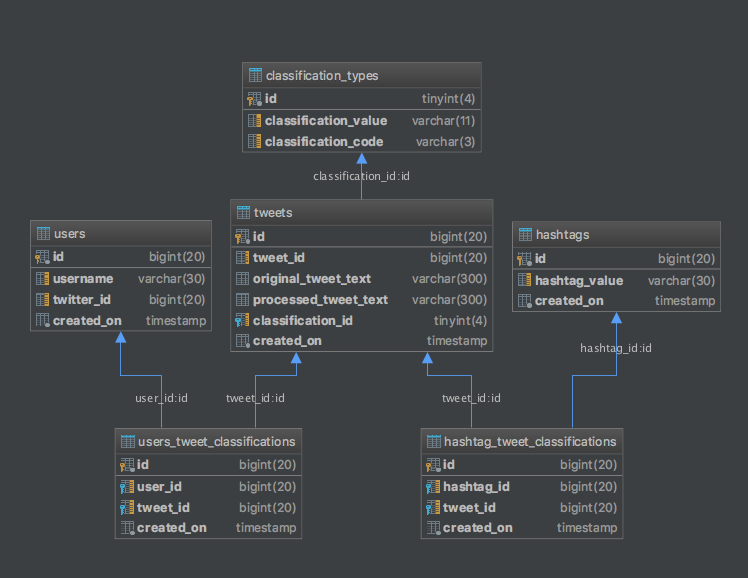
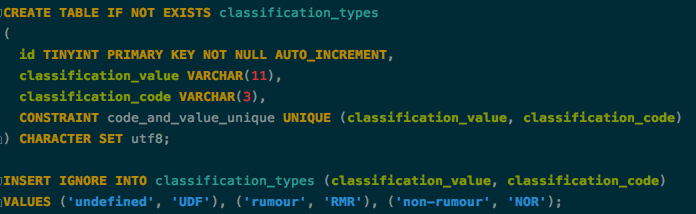
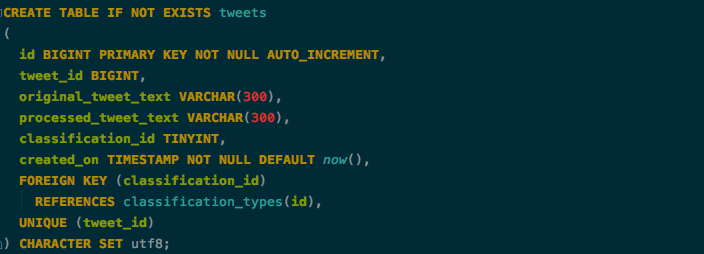


Figure 4 ER Diagram of Twitter\_Classification Database

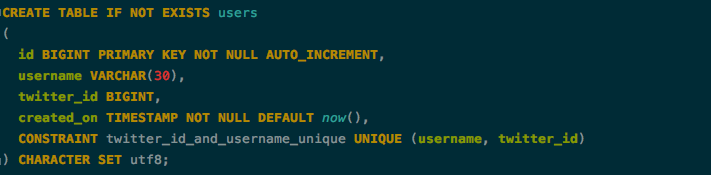
From figure 4 which illustrates the ER diagram of the database, the relations are indicated by the blue lines, i.e. table *users\_tweet\_classifications* will have a foreign key constraint on the table *users* through *users.id* to *users\_tweet\_classifications.user\_id*. With the ER Diagram in figure 4 it highlights how the database could be queried to get the Tweets and classification values for a particular hashtag or user of Twitter; which is important as this is the type of information which will be needed to present to the users of the service.



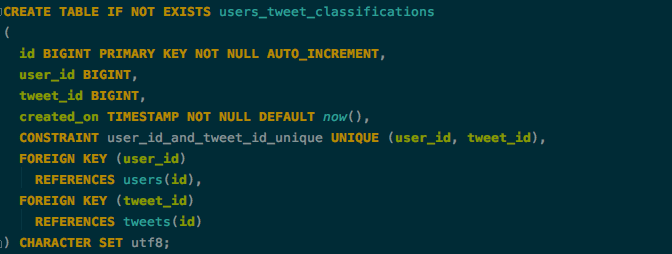
CODE 1 Create Table Statement for classification\_types



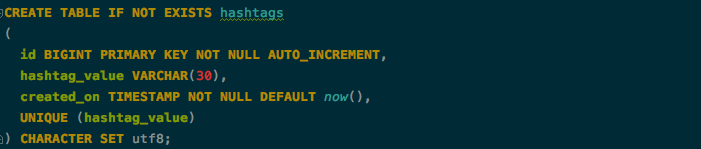
CODE 2 Create Table Statement for tweets



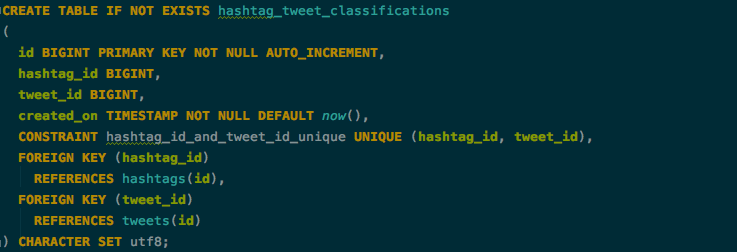
CODE 3 Create Table Statement for users



CODE 4 Create Table Statement for users\_tweet\_classifications



CODE 5 Create Table Statement for hashtags



CODE 6 Create Table Statement for hashtag\_tweet\_classifications

The code entries 1 through to 6 illustrates how the tables have been constructed along with the types of data each field accepts and the constraints which have been applied, code 1 is the statement which had been executed to create the *classification\_types* table which stores the two types of classification values which the system deals with; rumours and non-rumours, hence why it is created and then the *INSERT* statement is used to insert values for *undefined, rumour, and non-rumour –* undefined is a precautionary measure for instances which may have thrown an error during classification.

For the fields used, for *classification\_code* it is a field with a range of 3 characters as this is all the is required, *classification\_value* on the other hand has a range of 11 characters, which is one more than the largest character size of *non-rumour* and a constraint has been placed for the uniqueness of *classification\_code and classification\_value*, in order to ensure that there is only one entry for each classification type.

Code 2 illustrates the code executed to create the *tweets* table, which is where individual Tweets will be inserted, *original\_tweet\_text* is the original content of the Tweet without any processing done, and the *processed\_tweet\_text* is the text which went through the classification process and had processing applied, both of which have a field type of VARCHAR(300) which means they can only store 300 characters, and as the maximum size of a Tweet is 240 characters, this is enough space for them to be stored. From the code it indicates that there is a foreign key for *classification\_id* to *classification\_types.id* which also iterates what is indicated in the ER diagram on figure 4. As each *tweet* only gets assigned one classification the *tweet\_id* is unique in order to ensure that the same Tweet does not get processed more than once which is part of the data validation that is performed at the database level.

For the table *users* the creation is illustrated in code 3, and will store the *username* which is the screenname that the User of Twitter is known as, with a character limit of 30 set, while the combination of *twitter\_id* and *username* requires uniqueness to ensure that a particular users Tweets can be traced, as two users could have the same screen name, but two users will not have the same *twitter\_id and username*.

The *users* table is related to an individual Tweet through the *users\_tweet\_classifications* table represented in code 4 which has foreign keys to link the *user\_id­* to the *users.id* field and the *tweet\_id* to the *tweets.id* while the constraint is set to ensure that *user\_id* and *tweet\_id* is only stored once.

Code 5 is the code representing the creation of the *hashtags* table, which has a particular *hashtag\_value* which needs to be unique, as a hashtag on Twitter is a unique identifier it works well, similarly to *users* it has a field range of 30 characters, and again similarly it has a table *hashtag\_tweet\_classifications provided in code 6* to relate a particular *hashtag* to a particular *tweet* where the combination of *hashtag­\_id and tweet\_id* requires uniqueness with foreign keys to relate the *hashtag* to a *tweet* through *hashtags.id to hashtag\_tweet\_classifications.hashtag\_id* and then to relate to individual tweets through the *hashtag\_tweet\_classifications.tweet\_id* to the *tweets.id* relation.

## User Interaction Design

This section will provide the information about how users will interact with the system through the use of an Information Flow Diagram, which shows the exchange of information between system components (Fakhroutdinov K, 2018) which gives an indication of how the system will pass information from one entity to another at a high level, another method used to show the user interaction which is included in this section is the use of Use Case Diagrams, which provide an overview of the requirements of a system (UML 2 Use Case Diagrams: An Agile Introduction, 2018).

### Information Flow Diagram

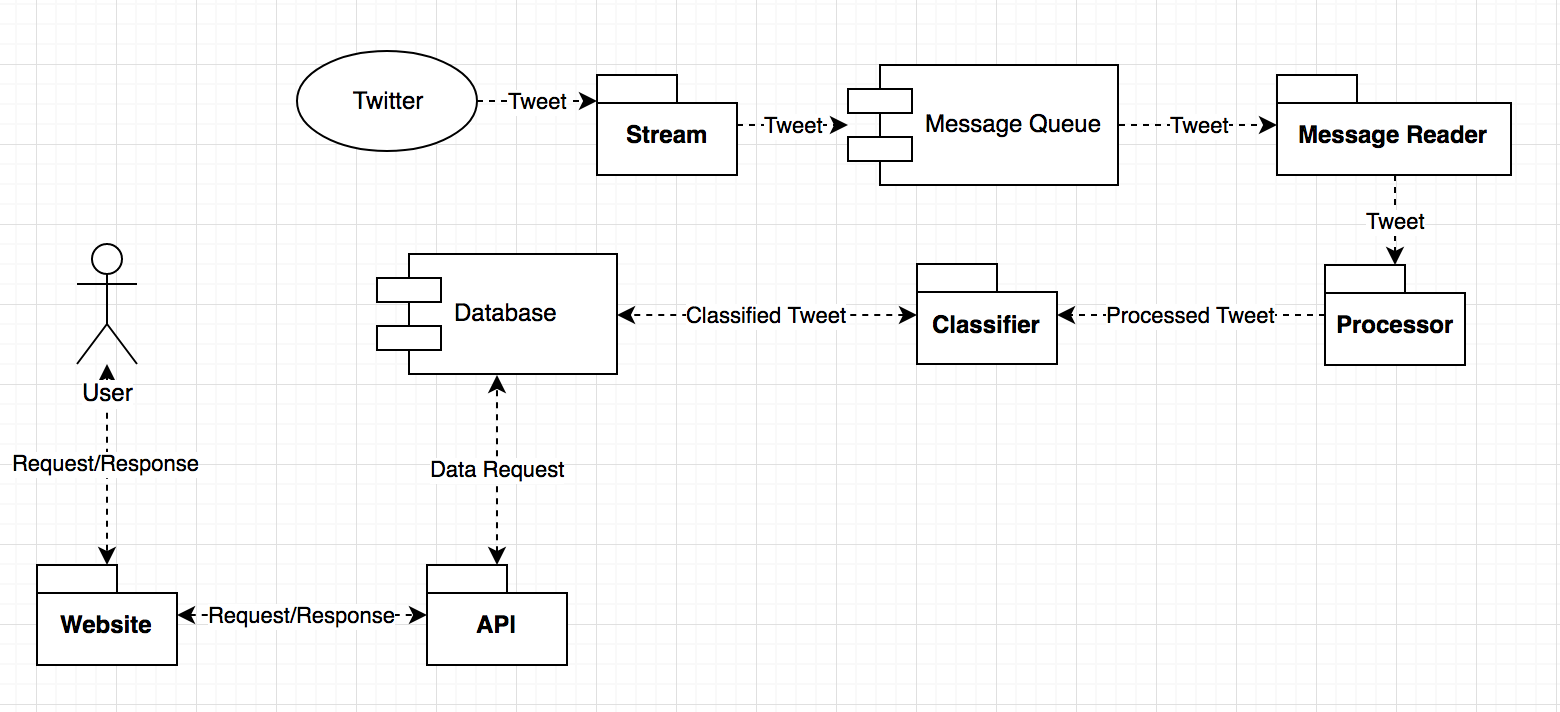


Figure 5 High-level Information Flow Diagram

Illustrated in figure 5 is the high-level information flow diagram for the project, where it shows the flow of data from the external service *Twitter* right down to the *Database –* and how the information is retrieved through data requests from the user. Starting at *Twitter*, the *Stream* retrieves an individual *Tweet* object which is then transferred on to the *Message Queue* – a *Message Reader* fetches the *Tweet* from this queue and dispatches the object to a *Processor* which transforms the original *Tweet* object in to a *Process Tweet* object which is then dispatched over to the *Classifier* which performs the classification task and creates a *Classified Tweet* object which is then stored in the *Database*. As referred to previously, this is the offline part which continuously works through incoming Tweets from Twitter, while on the left, the *User* makes requests to the website by navigating/searching sections which then makes a request call to the *API* which handles the request and fetches the data from the *Database*.

### Use Case Diagrams

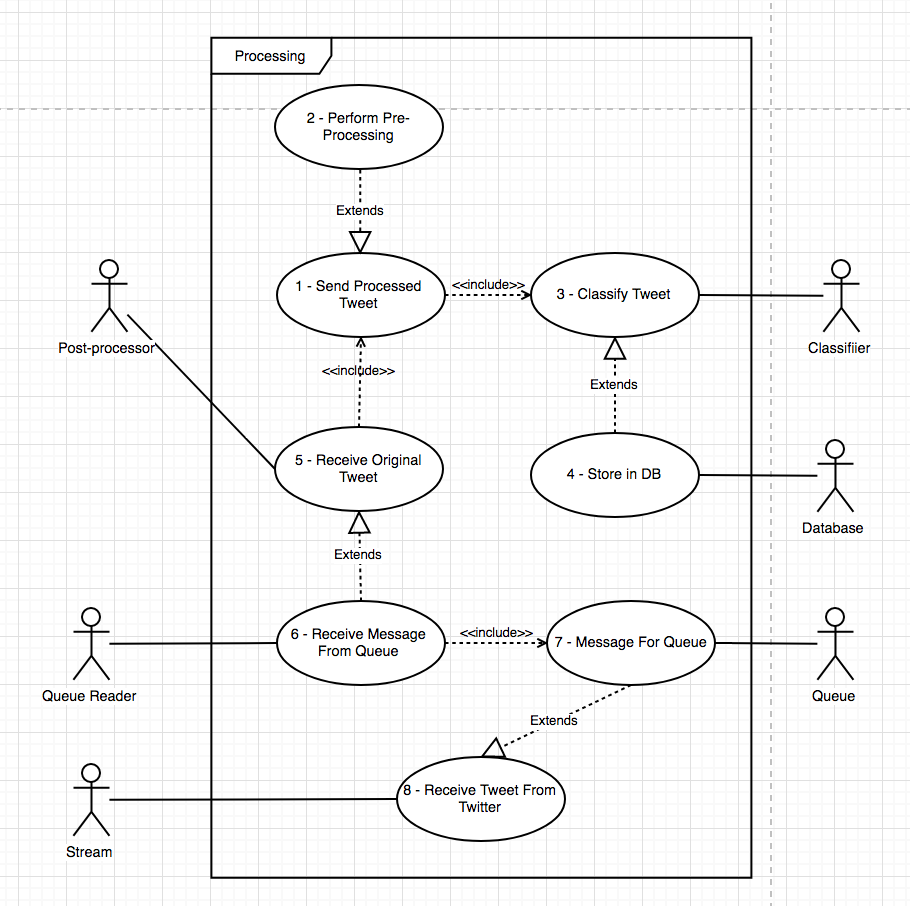


Figure 6 Use Case Diagram of the Offline Processing Part

A use case diagram is a visualisation of the set of actions which a system can perform in connection with users of the system (UML 2 Use Case Diagrams: An Agile Introduction. 2018) – users of the system are not necessarily end-users and can represent different entities in the system, such as the use case diagram provided in figure 6 where the users are different services of the project and it shows the relationships with different use cases (the actions).

With the use case diagram illustrated in figure 6, it shows the relationship between the entities in the offline part of the project, such as how the *Stream* relates to use case 8, receiving a Tweet from Twitter which is extended by the use case 7 – which stores it as a message to be stored in the *Queue*. As it represents the entities of the system with particular actions, it is a helpful way to cross check requirements and to ensure completeness.

### Activity Flow Diagram

An activity diagram is a flowchart to represent the flow from one activity to another activity, and usually described as an operation of the system (tutorialspoint.com, 2018), to visualise how a new Tweet will be processed on the system, figure 7 shows the activity diagram of a new Tweet entering the *Stream* services, and the flow of it through the system till it reaches the *Database*.

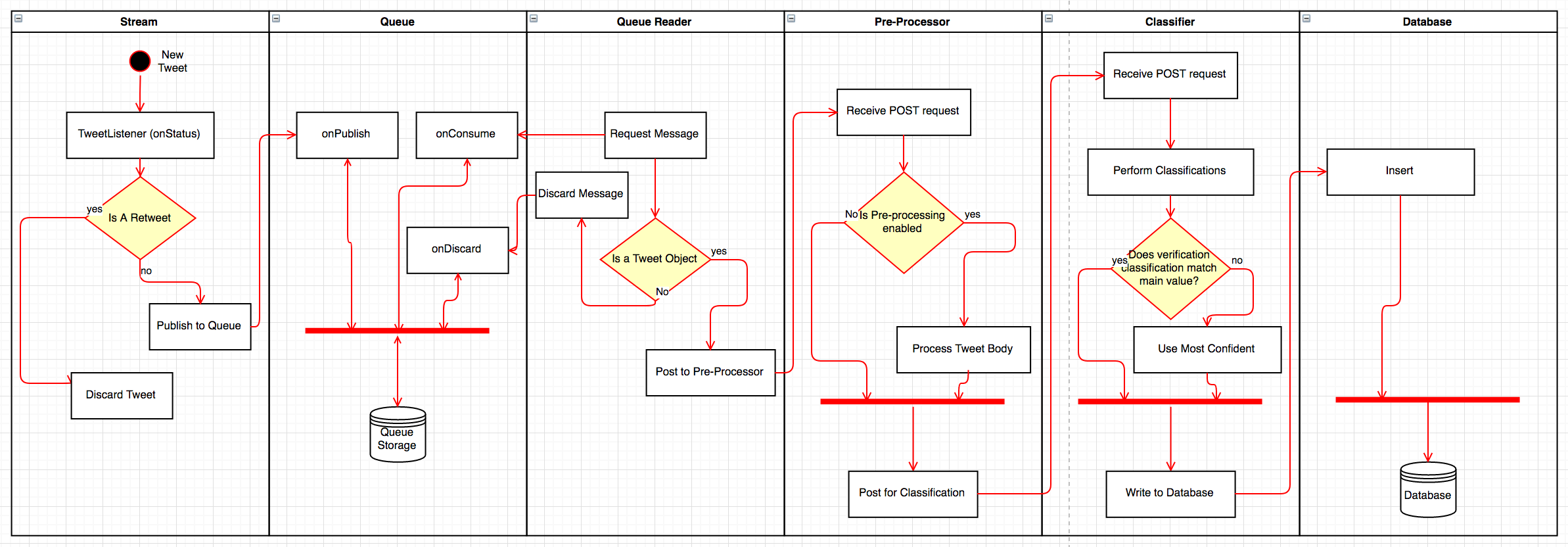


Figure 7 Activity Diagram of a 'New Tweet' entering the service

As illustrated in figure 7, a new Tweet will enter at the *Stream* which uses a Listener class to respond to new Tweets from the stream, from here it will check to see if the message is a Retweet, if it is – it’ll discard it as a Retweet would mean processing a Tweet which has already been done, if it is not a Retweet then the message can be added to the queue, which is done through publishing it, the queue will handle the storage of the Tweet.

The *Queue Reader* runs independently and has a consumer which handles new messages from the *Queue*, the *Queue Reader* will ensure the format of the message is correct and then it posts it to the *Pre-Processor* if the format is correct, otherwise it discards it and removes it from the *Queue*. Once the *Pre-Processor* receives a POST request, it will check to see if *Pre-processing* is required, which is enabled by a configuration variable, if it is, then it will perform the pre-processing steps outlined in section 4 and then it will POST the processed Tweet for classification, if the pre-processing variable is not set then it will POST the Tweet without any processing done.

The *Classifier* receives the POST request and performs the classification using the two trained classifiers, and then does a check to see if the result is the same, if it is it can write the results to the *Database*, if the results differ then it will go based on a *tolerance weight* set up as a configuration variable, if the weight of the main classification is below the tolerance level, then it will use the verification classification result, otherwise if the weight is above the tolerance then it will use the main classification value. Using an activity diagram to visualise the flow of events helped in the development of the project as it helped to ensure tasks were being executed when expected.

## Other Design Artefacts

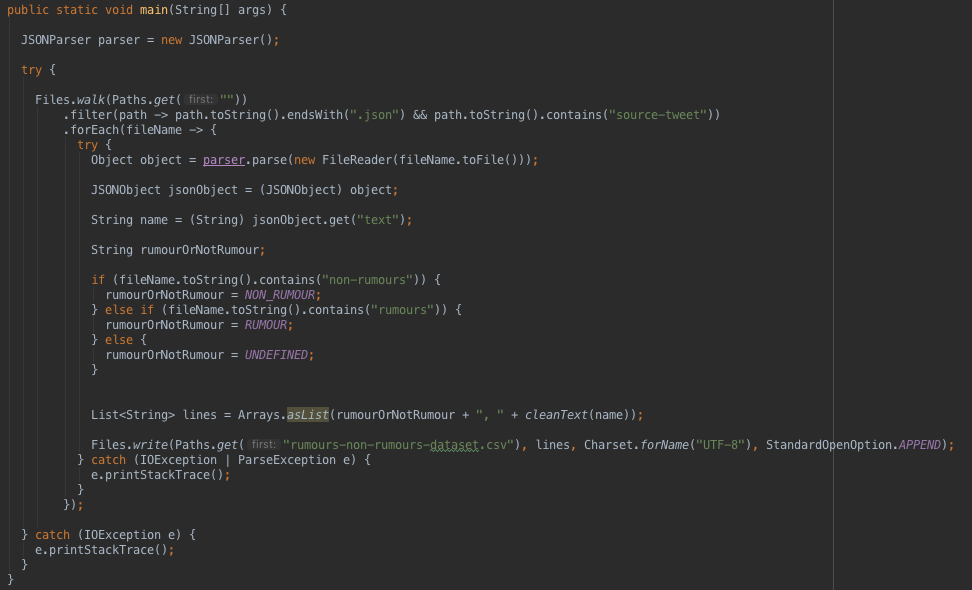
This section will outline the classification process which has been selected for the project, as the classification of a Tweet is the most important process a few considerations had to be made early to ensure that the project is as accurate as it can be.

### Classification Process

As the project will analyse Tweets and classify them as either a rumour, or non-rumour this process is known as a binary classification, which is the process of classifying given document on the basis of a predefined class (Srivastava S and Kumari R, 2017). In order to classify a new document, in this case a Tweet as a rumour or non-rumour there are a few preliminary steps, first of all a dataset is required which can be used for training a classification model.

The project went with a dataset which contained a collection of Twitter rumours and non-rumours posted during 5 breaking news events (PHEME, 2018); which provided a dataset containing ~3800 non-rumours and ~2000 rumours, although swayed more towards non-rumours it was the most ideal for what was required as the dataset was a collection of Tweets, and the new documents would also be Tweets.

Some steps had to be performed on the dataset to allow it to be used with the classification models, as the original structure of the data was a folder per event, and within each of the five folders were a rumours and non-rumours folder which contained the Tweets as a JSON object, in order to use these Tweets, they had to be split in to a CSV format as follows; *label, text* where the *label* will either be rumour or non-rumour and the text will be the Tweet text of the Tweet object, code listing 7 shows the code which created the CSV file;



CODE 7 Java Code to convert the PHEME dataset in to a CSV file

In code 7, it filters the files from (PHEME, 2018) dataset which end in ‘.json’ and they are within a ‘source-tweet’ folder, the contents of which are then parsed as a JSON object, in order for the ‘text’ of the Tweet to be retrieved, as the JSON object contains other information, such as ‘userId’ etc. then based on if the file is in the rumour or non-rumour folder the label will be assigned, then this information is appended to a CSV file which is then used as the dataset to train the classification model.

As the classification will be done on natural language, a classification model had to be selected, and in the field of natural language classification there is often three classification models which get considered; Naïve Bayes, Maximum Entropy, and Support Vector Machines (Go et al., 2009; Lee et al., 2011; Pang et al., 2002; Medlock B, 2008), early in the project plan it was swayed towards the use of Naïve Bayes as it is drastically more efficient in terms of training time and individual classification time.

Naïve Bayes is a classification model based on Bayes Theorem and it assumes that the presences of a particular feature is unrelated to the presence of another feature – hence the term naïve (Analytics Vidhya, 2018). As a result, it can calculate the class through what is known in Bayes’ Theorem as the posterior probability, illustrated in figure 8.

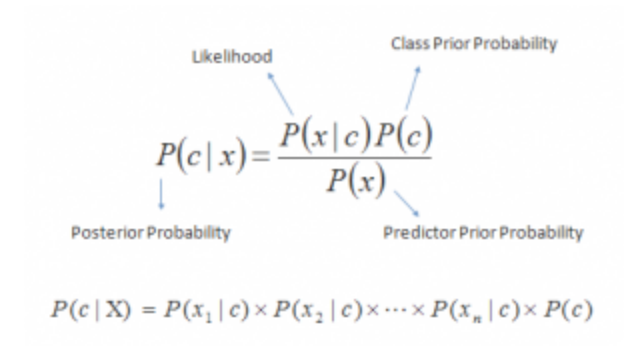


Figure 8 Bayes' Theorem Posterior Probability equation (Analytics Vidhya, 2018)

Which can be summarised as follows; given a set of attributes (*X*) the probability of the class (*c*) – *P(c|X)* where *X* is a set of attributes, in the case of natural language processing it would be text split in to a feature list, and *c* would be the class/label which is being evaluated is equal to the probability of x1…n belonging to class *c* multiplied by the class prior probability (*P(c)*) which is the maximum likelihood of something belonging to that class a rough example is as follows;

The same equation would be done for ‘non-rumour’ and the class the highest posterior probability value would be selected as the appropriate label for the text.

# Bibliography:

A

Analytics Vidhya. 2018. 6 Easy Steps to Learn Naive Bayes Algorithm (with code in Python) . [ONLINE] Available at: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/. [Accessed 10 April 2018].

Apache Tomcat Project. 2018. Apache Tomcat® - Welcome!. [ONLINE] Available at: http://tomcat.apache.org/. [Accessed 07 April 2018].

B

Ben Shneiderman. 1997. Designing the User Interface: Strategies for Effective Human-Computer Interaction (3rd ed.). Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Ben W. Medlock, Investigating classification for natural language processing tasks, Tech. Report UCAM-CL-TR-721, University of Cambridge, Computer Laboratory, June 2008.

Bloomberg.com. 2018. Fake News, Trump and the Pressure on Facebook: QuickTake Q&A - Bloomberg. [ONLINE] Available at: https://www.bloomberg.com/news/articles/2016-11-25/fake-news-trump-and-the-pressure-on-facebook-quicktake-q-a. [Accessed 07 April 2018].

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 (EMNLP '02), Vol. 10. Association for Computational Linguistics, 79-86 [ONLINE] Available at: http://delivery.acm.org/10.1145/1120000/1118704/p79- pang.pdf [Accessed 11 October 2017].

C

Checkmarx. 2018. The Importance of Database Security and Integrity. [ONLINE] Available at: https://www.checkmarx.com/2016/06/24/20160624the-importance-of-database-security-and-integrity/. [Accessed 09 April 2018].

Creately Blog. 2018. Use Case Diagram Relationships Explained with Examples - Creately Blog. [ONLINE] Available at: https://creately.com/blog/diagrams/use-case-diagram-relationships/. [Accessed 10 April 2018].

D

Database Trends and Applications. 2018. Bad Database Standards Can Cause Performance Problems - Database Trends and Applications . [ONLINE] Available at: http://www.dbta.com/Columns/DBA-Corner/Bad-Database-Standards-Can-Cause-Performance-Problems-91192.aspx. [Accessed 09 April 2018].

DigitalOcean. 2018. How To Create a New User and Grant Permissions in MySQL | DigitalOcean. [ONLINE] Available at: https://www.digitalocean.com/community/tutorials/how-to-create-a-new-user-and-grant-permissions-in-mysql. [Accessed 09 April 2018].

Docker. 2018. What is a Container | Docker. [ONLINE] Available at: https://www.docker.com/what-container. [Accessed 07 April 2018].

E

Experience UX. 2018. What is wireframing | Experience UX. [ONLINE] Available at: https://www.experienceux.co.uk/faqs/what-is-wireframing/. [Accessed 09 April 2018].

Extreme Programming: A Gentle Introduction.. 2018. Extreme Programming: A Gentle Introduction.. [ONLINE] Available at: http://www.extremeprogramming.org/. [Accessed 09 April 2018].

F

Facebook and Twitter unveil plans to tackle fake news | News | Ecommerce Week. 2018. Facebook and Twitter unveil plans to tackle fake news | News | Ecommerce Week. [ONLINE] Available at: http://www.ecommerceweek.co.uk/news/683/facebook-and-twitter-unveil-plans-to-tackle-fake-news/. [Accessed 07 April 2018].

Fake news definition and meaning | Collins English Dictionary. 2018. Fake news definition and meaning | Collins English Dictionary. [ONLINE] Available at: https://www.collinsdictionary.com/dictionary/english/fake-news. [Accessed 07 April 2018].

G

Go, Bhayani, Huang, 2009. Twitter Sentiment Classification using Distant Supervision. Final. Computer Science: Stanford University.

H

I

J

K

Karl Wiegers. 1999. First things first. Softw. Dev. 7, 9 (September 1999), 48-53.

Kate Conger. 2018. Here Are 14 Russian Ads That Ran on Facebook During The 2016 Election. [ONLINE] Available at: https://gizmodo.com/here-are-14-russian-ads-that-ran-on-facebook-during-the-1820052443. [Accessed 07 April 2018].

Kirill Fakhroutdinov. 2018. UML Information Flow Diagrams - Overview of Graphical Notation. [ONLINE] Available at: https://www.uml-diagrams.org/information-flow-diagrams.html. [Accessed 09 April 2018].

Kirill Fakhroutdinov. 2018. Use case diagrams are UML diagrams describing units of useful functionality (use cases) performed by a system in collaboration with external users (actors).. [ONLINE] Available at: https://www.uml-diagrams.org/use-case-diagrams.html. [Accessed 09 April 2018].

L

Lee K, Palsetia D, Narayanan R, Patwary M, Agrawal A, and Choudhary A, 2011. Twitter Trending Topic Classification. Final. Department of Electrical Engineering and Computer Science: Northwestern University.

M

MALLET homepage. 2018. MALLET homepage. [ONLINE] Available at: http://mallet.cs.umass.edu/. [Accessed 10 April 2018].

Meyer, Robinson. 2018. Huge MIT Study of ‘Fake News’: Falsehoods Win on Twitter - The Atlantic. [ONLINE] Available at: https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/. [Accessed 07 April 2018].

MySQL :: MySQL 5.7 Reference Manual :: 6.3.5 Setting Account Resource Limits. 2018. MySQL :: MySQL 5.7 Reference Manual :: 6.3.5 Setting Account Resource Limits. [ONLINE] Available at: https://dev.mysql.com/doc/refman/5.7/en/user-resources.html. [Accessed 09 April 2018].

MySQL :: MySQL 5.7 Reference Manual :: 13.7.1.4 GRANT Syntax. 2018. MySQL :: MySQL 5.7 Reference Manual :: 13.7.1.4 GRANT Syntax. [ONLINE] Available at: https://dev.mysql.com/doc/refman/5.7/en/grant.html. [Accessed 09 April 2018].

MySQL :: MySQL 5.7 Reference Manual :: 13.5 Prepared SQL Statement Syntax. 2018. MySQL :: MySQL 5.7 Reference Manual :: 13.5 Prepared SQL Statement Syntax. [ONLINE] Available at: https://dev.mysql.com/doc/refman/5.7/en/sql-syntax-prepared-statements.html. [Accessed 09 April 2018].

N

Nick Babich. 2018. Golden Rules of User Interface Design. [ONLINE] Available at: http://babich.biz/golden-rules-of-user-interface-design/. [Accessed 09 April 2018].

Nielsen Norman Group. 2018. 10 Heuristics for User Interface Design: Article by Jakob Nielsen. [ONLINE] Available at: https://www.nngroup.com/articles/ten-usability-heuristics/. [Accessed 09 April 2018].

NGINX. 2018. Microservices at Netflix: Lessons for Architectural Design. [ONLINE] Available at: https://www.nginx.com/blog/microservices-at-netflix-architectural-best-practices/. [Accessed 07 April 2018].

O

P

PHEME. 2018. PHEME dataset of rumours and non-rumours. [ONLINE] Available at: https://figshare.com/articles/PHEME\_dataset\_of\_rumours\_and\_non-rumours/4010619. [Accessed 10 April 2018].

POST statuses/filter — Twitter Developers. 2018. POST statuses/filter — Twitter Developers. [ONLINE] Available at: https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html. [Accessed 07 April 2018].

Q

R

S

Science | AAAS. 2018. Fake news spreads faster than true news on Twitter—thanks to people, not bots | Science | AAAS. [ONLINE] Available at: http://www.sciencemag.org/news/2018/03/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots. [Accessed 07 April 2018].

SearchDataManagement. 2018. Three indicators that could signal database performance issues. [ONLINE] Available at: https://searchdatamanagement.techtarget.com/feature/Three-indicators-that-could-signal-database-performance-issues. [Accessed 09 April 2018].

SearchMicroservices. 2018. What is RESTful API? - Definition from WhatIs.com. [ONLINE] Available at: https://searchmicroservices.techtarget.com/definition/RESTful-API. [Accessed 07 April 2018].

Shneiderman's "Eight Golden Rules of Interface Design" | Design Principles FTW. 2018. Shneiderman's "Eight Golden Rules of Interface Design" | Design Principles FTW. [ONLINE] Available at: https://www.designprinciplesftw.com/collections/shneidermans-eight-golden-rules-of-interface-design. [Accessed 09 April 2018].

Srivastava S, Kumari R, 2017. Machine Learning: A Review on Binary Classification. Internation Journal of Computer Applications, [Online]. Vol. 160, No. 7, 11-15. Available at: https://pdfs.semanticscholar.org/55b0/b0478fad2cd6c70d851e2be594a2788910ff.pdf [Accessed 10 April 2018].

T

The Guardian. 2018. New Facebook controls aim to regulate political ads and fight fake news | Technology | The Guardian. [ONLINE] Available at: https://www.theguardian.com/technology/2018/apr/06/facebook-launches-controls-regulate-ads-publishers. [Accessed 07 April 2018].

tutorialspoint.com. 2018. UML - Activity Diagrams. [ONLINE] Available at: https://www.tutorialspoint.com/uml/uml\_activity\_diagram.htm. [Accessed 10 April 2018].

Twitter4J - A Java library for the Twitter API. 2018. Twitter4J - A Java library for the Twitter API. [ONLINE] Available at: http://twitter4j.org/en/index.html. [Accessed 07 April 2018].

U

UML 2 Use Case Diagrams: An Agile Introduction. 2018. UML 2 Use Case Diagrams: An Agile Introduction. [ONLINE] Available at: http://www.agilemodeling.com/artifacts/useCaseDiagram.htm. [Accessed 09 April 2018].

UX courses. 2018. What is Human-Computer Interaction (HCI)? | Interaction Design Foundation. [ONLINE] Available at: https://www.interaction-design.org/literature/topics/human-computer-interaction. [Accessed 09 April 2018].

UX courses. 2018. User Interface Design Guidelines: 10 Rules of Thumb | Interaction Design Foundation. [ONLINE] Available at: https://www.interaction-design.org/literature/article/user-interface-design-guidelines-10-rules-of-thumb. [Accessed 09 April 2018].

UX courses. 2018. Shneiderman’s Eight Golden Rules Will Help You Design Better Interfaces | Interaction Design Foundation. [ONLINE] Available at: https://www.interaction-design.org/literature/article/shneiderman-s-eight-golden-rules-will-help-you-design-better-interfaces. [Accessed 09 April 2018].

V

Validation | Databases | ICT. 2018. Validation | Databases | ICT. [ONLINE] Available at: http://www.advanced-ict.info/databases/validation.html. [Accessed 09 April 2018].

Vox. 2018. Russia, Donald Trump, and “fake news”: How Russia pioneered the art of fake news - Vox. [ONLINE] Available at: https://www.vox.com/world/2018/4/5/17172754/russia-fake-news-trump-america-timothy-snyder. [Accessed 07 April 2018].

W

Weka 3 - Data Mining with Open Source Machine Learning Software in Java . 2018. Weka 3 - Data Mining with Open Source Machine Learning Software in Java . [ONLINE] Available at: https://www.cs.waikato.ac.nz/ml/weka/. [Accessed 07 April 2018].

X

Y

Z

1. false, often sensational, information disseminated under the guise of news reporting (Fake news definition and meaning | Collins English Dictionary, 2018) [↑](#footnote-ref-1)
2. Loosely coupled means that you can update the services independently; updating one service doesn’t require changing any other services. (NGINX, 2018) [↑](#footnote-ref-2)
3. Human Computer Interaction - a field of study focusing on the design of computer technology and, in particular, the interaction between humans (the users) and computers (UX Courses, 2018) [↑](#footnote-ref-3)
4. <http://twitter4j.org/en/index.html> – Twitter4J is an unofficial Java library for the Twitter API. (Twitter4J, 2018) [↑](#footnote-ref-4)
5. <https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html> - Returns public statuses that match one or more filter predicates. (POST statuses/filter — Twitter Developers. 2018) [↑](#footnote-ref-5)
6. A RESTful API is an application program interface (API) that uses HTTP requests to GET, PUT, POST and DELETE data. (SearchMicroservices, 2018) [↑](#footnote-ref-6)
7. The Apache Tomcat® software is an open source implementation of the Java Servlet, JavaServer Pages, Java Expression Language and Java WebSocket technologies. (Apache Tomcat Project, 2018) [↑](#footnote-ref-7)
8. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. (Weka 3 - Data Mining with Open Source Machine Learning Software in Java, 2018) [↑](#footnote-ref-8)
9. MALLET is a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text. (MALLET, 2018) [↑](#footnote-ref-9)