# MOBILE PHONE SALE PREDICTOR USING SOCIAL MEDIA DATA FROM YOUTUBE

LOH JIA JING



UNIVERSITI TEKNIKAL MALAYSIA MALAKA

BORANG PENGESAHAN STATUS TESIS

JUDUL : **MOBILE PHONE SALE PREDICTOR USING SOCIAL MEDIA DATA FROM YOUTUBE**

SESI PENGAJIAN : **2015 / 2016**

SAYA **LOH JIA JING**

(HURUF BESAR)

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah) ini disimpan di perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dengan syarat-syarat kegunaan seperti berikut :

1. Tesis dan projeck adalah hakmilok Universiti Teknikal Melaka
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara insitusi pengajian tinggi.
4. **Slia tindakan (/)

| | | |
|---|---|---|
| _____ | SULIT | (Mengandingi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termektub di dalam AKTA RAHSIA RASMI 1972) |
| _____ | TERHAD | (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan) |
| / | | |
| _____ | TIDAK TERHAD | |



(TANDATANGAN PENULIS)

Alamat tetap: 801, MK 20
Kubang Semang, 14400
Bukit Mertajam, Pulau Pinang

Tarikh : 26/8/2016

(TANDATANGAN PENYELIA)

Assoc. Prof. Dr Choo Yun Huoy

Nama Penyelia

Tarikh : 26/8/2016

CATATAN: * Tesis dimaksudkan sebagai Laporan Akhir Projek Sarjana Muda (PSM)
** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada phak berkuasa

**MOBILE PHONE SALE PREDICTOR USING SOCIAL MEDIA DATA FROM YOUTUBE**

LOH JIA JING



This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Artificial Intelligence)

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA
2015

**DECLARATION**

I hereby declare that this project report entitled

**MOBILE PHONE SALE PREDICTOR USING SOCIAL MEDIA DATA FROM YOUTUBE**

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ DATE : 26/8/2016

(LOH JIA JING)



I hereby declare that I have read this project report and found this project report is sufficient in term of the scope and quality for the award of Bachelor of Computer Science (Artificial Intelligent)

SUPERVISOR : _____ DATE : 26/8/2016

(ASSOC. PROF. DR CHOO YUN HUOY)

**DEDICATION**

To my beloved family, especially my parents, Loh Siak Kee and Teoh Ean Teang.

To my respected supervisor, Associate Professor Dr Choo Yun Huoy.

To my dearest classmate, girlfriend and life partner, Tey Woan Jiuan, Lim Chun Heng and Raymond Lee Wei Kiat.

## ACKNOWLEDGEMENTS

I would like to show my highest appreciation to my parents for always supporting me in term of spiritual and financial. Although they are not around, but their love will always by my side to make me move forward.

Secondly, I want to thank to my supervisor, Associate Professor Dr Choo Yun Huoy. Besides of supervising me, she also show great support and help me to solve the problems. Dr Choo is so knowledgeable and specialist in data analysis, with this she always give suitable guidance to me throughout my project.

## ABSTRACT

The basis of this project is to find out the impact of social media data on sale forecasting. Leveraging the power of YouTuber on tech reviewing to forecast the sale of Apple iPhones. First, the media data like views counts, number of dislike and like on the video about iPhones are collected through YouTube API. System will then transform these unstructured data to useful informations. Sentiment analysis is also carried out to determine the positivity of user's comment. Combine all the information collected and visualize on interactive real-time dashboard are also one of the features of the system. In the end, we will use the historical sale report of the devices and the information to generate a best model to forecast the sale.

# ABSTRACK

Tujuan projek ini adalah mengenal pasti kesan data media sosial terhadap ramalan jualan. Dengan memanfaatkan kelebihan YouTube untuk meramal jualan Apple iPhones. Data media seperti bilangan tontonan, bilangan suka, bilangan tidak suka video-video tentang iPhones yang dikumpul dengan mengunakan YouTube API. Kemudianya, sistem itu akan mengubahkan data-data yang tidak berstruktur ini kepada maklumat yang bernilai. Analisis Sentimen juga dijalankan untuk menentukan nilai positif komen terhadap video-video tersebut. Sistem ini kemudian menggabungkan semua informasi yang dikumpul dan menggambarkannya di atas laman interaktif masa nyata. Akhirnya, dengan mengunakan sejarah laporan jualan dan informasi-informasi untuk menjana model ramalan.

# TABLE OF CONTENTS

**CHAPTER II LITERATURE REVIEW AND PROJECT METHODOLOGY**





**CHAPTER III ANALYSIS**

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| API | - | Application Programming Interface |
| ANOVA | - | Analysis of Variance |
| ANCOVA | - | Analysis of Covariance |
| SDLC | - | Software Development Life Cycle |
| OS | - | Operating System |
| JSON | - | JavaScript Object Notation |
| AI | - | Artificial Intelligence |
| HTTP | - | Hypertext Transfer Protocol |
| UTeM | - | Universiti Teknikal Melaka Malaysia |

# CHAPTER I

# INTRODUCTION

## 1.1 Introduction

Nowadays, Technology grow exponentially. Following Moore's law, computer's processor increasing with double-up speed every 18 months. It is proven that this law has already broke, as the tech today is growing even aggressive than that, especially in mobile phone field. Report showed that mobile phone's sale hits 2 billions in past few years and the growth still continue strongly. Every year, the big mobile tech giants, like Apple, Samsung, Huawei, Google and others more, released more than 20 kinds of new model of cell phones.

In this century of internet, social networks are playing a big role in connecting people where people can share their opinion to their friends around the worlds. Social Media is what have been shared by the user, it can be in form of text, image, animation or videos on the social networks. Here is some big example of famous social networks, like Facebook, Twitter, Instagram and Tumblr. The users of social network have no limited to age, far from the primary school students to the 65 years old are contributing themselves to social network everyday. Social Media not only represent the opinions of individual to share to the world. Back to the business part of the world, social network can be a main platform to advertise and promote the

companies' products. Social Media in this term can be represent the important informations that being produced to establish the connection between the consumer and producer. What we said that, the feedback toward a product in social network is solid and direct. Social Network is free of talk, any kind of languages and way of expressing are all acceptable on Social Network. As the result, tons and overwhelming of social data are generated everyday. This is extremely essential in business side as for the company to understand their user more by analysis and transform those data into useful and valuable informations.

YouTube has become the most widely used video sharing platform. Report declared that there are more than 4 billions of video viewed each day and more than 1 million videos is uploaded to YouTube every single day. This rapid growth of number of users as well as videos happening, thanks to the simplicity usage of the system and the interesting contents. The feedback on YouTube is very direct. Users can choose to click on the dislike or like buttons or leave the comments under the video. There are no limitations of category of video on YouTube, which mean any kind of contents can be founds, which including sport, gaming, musician, movie, tech review, culinary and many more.



Let's focus on tech review. There is not hard to search for the tech review videos on YouTube. Basically what tech review does is explaining the latest technology or electronic gadget and sharing their opinions after using those particular items. These tech reviewers share their feeling and give back feedback to all the people after trying the new devices for few days. This review videos bring to the win-win situation to both consumers and producers. Consumers can always understand whether the new devices is fulfilling their need and make a deep consideration before purchasing the product. In the order hands, producer can always know how good is the product they just released and directly understand what is actually needed by consumers. The famous YouTube Tech review Channel, for example MKBHD, The Verge and TechnoBuffalo which can attract more than 100

thousand views in single video and make over 20% of user that viewed the video to leave their comment under the video.

Sales forecasting is important in helping a business to grow. As we know, cash is playing a big role in business as it provides a working capital. For example, cash need to pay to vendor or manufacturer before the products are published to market. Consequently, business need to prepare forecast so they can predict when problems are likely to occur. Sale forecasting ease the process of handling of cash flow. Traditional way of research is one of the method to forecast the sale. However, social media monitoring come to the market and it is considered less expensive than traditional market research. Not only that, social media monitoring also reflected the real feedbacks of customer.

One of the reason we choose YouTube Data is to lover down the threshold of forecasting. Google provides a powerful application programming interface(API) to crawl their public data. This API is totally free to use, but there are some limitation of it which concluded in the chapter later. In overall, it is useful and well-documented especially when it come to accessibility for public and useful data for analytic purpose.



In this project, we will be leveraging the power of YouTuber on tech reviewing to forecast the sale on the mobile phones. By collecting the media data like views counts, number of dislike and like on the video and transform these unstructured data to useful information. Sentiment analysis is also carried out to determine the positivity of user's comment. Combine all the information collected and visualize on interactive real-time dashboard are also one of the features of the product. In the end, we will use the historical sale report of the devices and the information to generate a best model to forecast the sale.

## 1.2 Problem Statement

There are many research on analysis the social media data especially twitter data. Twitter data provides not more than 140 words of text and hashtag which is easier to process. Based on the number of text or hashtag mentioned by the users, the next trends can be predicted. However, as the aggressive increasing of number of users and videos on YouTube, the YouTube data now becoming one of the essential information to reveal people's perspective on one product. The problem is that many researchers didn't focus much on YouTube data. Social Media monitoring is considered less expensive than traditional market research. And It reflected the real feedback of customer toward one products. However, the cost to collect and analysis data might be high. Therefore, in this project, we will be using the YouTube data to forecast the number of sales of mobiles devices in the future.

## 1.3 Objective

1. To design a real-time application using the Meteor framework

   Meteor is a JavaScript framework made for building real-time web application. A real time application can allow user to latest information and view the update on the dashboard with refreshing their browser manually. Therefore, the user will always up-to-date.

2. To visualize the data collected from Youtube on the interactive dashboard. Instead of using text form, all the data will be displayed in the colorful and interactive dashboard which is made by bar charts, table, graph, pie chart and bar columns. In this way, user can immediately understand the content without deep thinking.

3. To develop the suitable algorithm on running a prediction on the mobile phone.

   As we all the unstructured data such as view counts, dislike counts, like counts, and positivity of the comments, we need to find out the best model to predict the sales.

4. To asses the usefulness of YouTube data on business sale predictions.

   In the end, we need to test whether our algorithm is accurate. The prediction result need to be accuracy enough to bring impact to the business.

**1.4 Scope**

This project focuses on three main modules, which are YouTube API, Intelligence Module and System Module.

1. YouTube API

- There are 4 most common methods that the API supports which are list, insert, update and delete. List method is to retrieve (GET) resources; Insert method is for creating (POST) a new resources. Update method is to modify (PUT) an existing resources. Lastly, delete method is to remove (DELETE) a specific resource. In this project, the method used is only List method.

- As the only method used is List, the system does not require user authorization and API key method is selected over OAuth 2.0 authorization.

- Quota usage is calculated by assigning a cost to each request. The GET request has a cost of 1 unit per request.

- The total quota per day is limited to 50,000,000 which is around 10000000 read operations per day as well as 300,000 unites per 100 seconds per user.

- The total number of comments that can be retrieved is only limited to 50% of each original amount. For example, there are 4000 comments made under one particular video. Therefore only 2000 comments can be requested and retrieved to the system.

- The search resources is obtain with GET request to return a collection of search results that match the query parameters specified in the request.

- The order parameter is set to relevance, where API will return relevance items that match the query parameters. Therefore, the videos retrieved might different than what expected. For example, if "iPhones 1" is search, the return of resources might contain videos of "iPhone 6". This error is removed manually.

2. Intelligence Module

- Sentiment Analysis can only read language English. Other language will be ignored and set to score of 0 and positivity to neutral.

- Sentiment Analysis is performed by using AFINN-111 wordlist which is the latest version and contain 2477 words and phrases.

- The mobile phone that used to forecast is limited to Apple iPhone.
- The historical sale data of all iPhone from year 2007 to 2015 are retrieved manually from Apple Annual Report.

- Five features which are view count, like count, dislike count, number of positivity of comments and total score of comments are used to train the

model for forecasting by linear regression.

3. System Module

- The system is set to retrieve only 50 videos per phone. For example, there will be 50 videos about iPhone 6 and its data respectively.

- The system required 15 - 20 minutes to retrieve all data of that 50 videos.

- The system will remain unstable and lagging when request is sent until the all resources are return.

- The system is not fully responsive, might not be viewed on smaller screen.



## 1.5 Project Significance

The project significance can de divided into 2 different category, which are industrial significance and consumer significance.

First of all, the main significance is dropped on industrial area. For industrial, sale is like a heart of the business and cash can be represent the bloodstream. Cash is generated by sale to pay the salary to employee, cover operating expenses, purchase more inventory, market new products and attract more investor. A sales forecast is an important tool for managing a business. An accurate forecasting of sale can help to avoid unforeseen cash flow problems which will lead to failure of business in the end of the day. There are some benefit of having a good sale forecasting.

1. Demand handle

   With sale forecast, a business can easily understand and estimate the demand of the product. The sales team can get ready and into the position to gather

information about anticipated demand.

2. Inventory Control

   Inventory Control is important in one business. Overstock can cause of loss in profit for wasting the cost of manufacturing whereas stock-out can lead to loss of customer buying experiment or even loss of customer. With good sale forecast, these two situation can be resolved easily.

3. Supply Chain Management

   With predicting the demand, company can also have better supply chain control. Every products are made up by smaller components supplied by other industry. With this, over purchasing of resources can be avoided. Company will have better resources management and take full advantage on ordering

4. Marketing

   If sale is predicted to be go weak. The marketing team can re-schedule the promotions and propose the new strategy of marketing.



It is always wise to expect the uncertainty might happened. An accurate sale forecasting, can always allow the company to generate a better sale plan which can benefit the company. On the other hand, the consumer of the product can review the dashboard of the product to know how good is the product based on the comment made by others. They can also find out what videos is making a greatest impact to other consumer. By this, they will be safer to make a good consideration before purchasing the new devices.

## 1.6 Expected Result

The project is expected to predict the sales of iPhone 6s which going to announce end of this year. All the informations retrieved from YouTube API will be visualized nicely in the interactive dashboard. The system will be running real-time where user will always get the latest information from YouTube.

## 1.7 Conclusion

As the small summary of this chapter, the proposed application will be developed to analyze the media data from Social Network - YouTube and forecast the sales of the products which is iPhone. The result of the project is believed to bring a huge impact to the business. The YouTube data is also proposed to understand the valuable of media data as call as unstructured data. In the era of big data, tons of data are generated every single seconds. It's time to discover the usage of these data.

# CHAPTER II

# LITERATURE REVIEW AND PROJECT METHODOLOGY

## 2.1 Introduction

In order to complete this project within a defined schedule and budget, a lots of research are being done to clarify every step we take in the future are leading us to success. Every reason behind the step we moved must be justify with strong fact as the supports. Other than that a good and suitable methodologies must be selected to help us at every stage of a project from initiation to implementation to the closure.



A methodology is a model which can always guide us to follow the plan we scheduled. The model are mainly involved in 4 phase which are planning, designing, implementation and developing. There are many project management methodologies out there. However, in this project, waterfall development model is chosen. The project progresses to an orderly sequence of steps in waterfall development. From the initial software concept, down until the final phase. At the end of each phase, there will be a review to determine the project is ready to advance to the next phase. For detailed about methodologies, will be concluded in the sub-chapter below.

Literature reviewing is set as the first phase of this project. Many research paper is studied to ensure that the project is runnable. A lots of uncertainty are come to

resolve easily when more resource we gained during this phase. Beside this, we can also avoid failure of project by understanding the problem other researcher faced. Avoiding the mistake that taken by other also lead us to the bright path.

In general, this project consists of three important modules, which are Domain, Technology and Intelligence. Each modules will not affect the other modules, if it is failed working. Each modules can worked independently. However, to make sure theses three modules can work together perfectly, we defined everything in first phase. In Domain, we explained the reason behind we selecting YouTube, but not other social network. Regarding Technology modules, Meteor.Js is selected as the main development tools and we will justify the decision later. Lastly, which is on intelligence part, the best technique is selected after the research.

## 2.2 Social Media Marketing

Social media marketing have becoming the new trend of marketing in the business side due to the problems of money and time consuming on traditional market research. Compared to traditional method, social media marketing is cost-effective and manage to gain the insights into the business's customer, market, brand appearance and more important aspects, based on the journal reported by Ray Nelson(2013) on SocialMediaToday. As for this, gaining the traffic and attention of customer through social media sites has become the biggest challenge in the business site. There are so many social network platform. For example, Twitter, the social platform that allow users to share their thought in 144 words of short messages. YouTube is the world largest video sharing platform which allow billions of people to discover, watch and share the originally videos. Facebook, the world largest social network which contain any kind of social data in any kind of form, such as text, articles, images, videos, etc. Different strategy will be needed to response to different users on different social platform.

Does social media affect customer affect consumer decision making ? After the survey of Charles-Henri Gros (2012), it is clearly to see that 49% of consumer will purchased on the product after reading the positive review on social media against 14% after reading the negative reviews. Impact of positive comments reviews and comments of strangers on social media can boost the confidence of purchasing decision, and there are 46% of consumer can deny this truth. Erik Weber also reported that up to 53% of consumers are most probably influenced by YouTube reviews.

The study of Charles-Henri Gros (2012) also revealed the usage of consumer on multiple social platform to obtain information about products or services they want. 46% percents of consumer selected multimedia sharing platform like YouTube; whereas only 13% of consumers likely to use microblogging platform which is Twitter. Jess Mawhinney on the article of "Where Market Go to Grow" (Jan 2016) mentioned that people are more willing to read or view on colored visuals over non-color content by 80%. 64% of business who use video as marketing method believe that it led to increased sales.



There are some research are made on top of media data. Ce´dric Richier, Eitan Altman, Rachid Elazouzi, Tania Jimenez, Georges Linares and Yonathan Portilla using the view count of videos in YouTube to study empirically the impact of videos' popularity. In concordance with previous research, Sahar Nassirpour, Parnian Zargham and Reza Nasiri Mahalati using the Twitter data like comment and hash tag to predict the sale of electronic devices. Trend prediction seem to be common to be seem in the research paper. For example, Rong Lu, Zhiheng Xu, Yang Zhang and Qing Yang predict the trends of topics on Twitter using MACD index on stocks.

## 2.3 Sentiment Analysis and Predictions using Social data

There are many researchers have performed sentiment analysis of social networks such as Twitter and YouTube. Sentiment analysis can work perfectly to analysis the text based content, for example comments, tweets and other metadata. The main interesting point of media data is all data are fully generated by mass of users on social network which believed can help to find the insight about the usage of this social network. The last part of literature review will be focusing choosing predictive algorithm to build a model which suit the data collected. The most common way of analysis the positivity of comment is using sentiment analysis. This method is widely use in trend predicting application regardless the social data retrieve source.

Below are some existing system using the similar technique.

- **A Survey of Prediction using Social Media**

  In this research, Shang Yu and Subhash Kak discussed multiple prediction methods that can be used with social data. The first methods suggested is Regression methods, which use to analyze relationship between every dependent and prediction result. The result also showed that regression models does not bring the big impact when working with sentiment analysis in movie. However, regression is the simplest and most used prediction method. Next, they talked about Bayes classifier, which using the formula to calculate the probability, that the object belongs to the result class. The class with largest posterior probability is selected, because the new event will more likely to happen. They added that Bayes classifier can be used directly to the system if the prediction result is discrete. Sheng Yu and Subhash Kak also talked more on Machine Learning prediction method which are K-nearest Neighbor classifier, Artificial Neural Network, Decision Tree and Model Based Prediction. However, the accuracy of the prediction result and the effective of the particular methods are not defined in this researched.

- **Product Sales Prediction Based on Sentiment Analysis Using Twitter Data**

  Generally, what Dipak Gaikar and Bijith Marakarkandy did in this research is to predict the movie sales based on the tweet made on twitter. The twitter data will first extracted and stored in the database by Twitter API. Sentiment Analysis is run to determine whether the message extracted is positive, negative, strongly positive or strongly negative. They believed that the result of the research will help in business to solve and meet the expectation of consumers. The prediction method was performed using ANOVA, ANCOVA and Multiple regression after the data collection.

- **Predicting the Future with Social Media**

  With no exception, this research utilize the power of social media content to predict real-world outcome. For instances, Sitaram Asur and Bernardo A. Huberman use the chatter from Twitter to predict the box-office revenues for movies. They also added, it is very common to label the text as Positive, Negative or Neutral by sentiment analysis. As for the research, they tried a new method which constructed a sentiment analysis using LingPipe linguistic analysis package which open source java library for natural language processing task is provided. The coefficient of determination is used to create the regression model. The result showed that simple linear regression model was good enough to run the prediction and revealed the secret of social media.



- **Polarity Trend analysis of Public Sentiment on YouTube**

  In this project, 6000 videos YouTube is analyzed and more than 7 million comments is collected to perform sentiment and prediction analysis. Amar Krishna use the Naive Bayes classification to find out the polarity of each comment. Due to the limitations of YouTube API, Amar only able to extract maximum of 1000 comments per video. To simplify the process of sentiment analysis, pre-process of dataset was done. The other language then English comment and all the emoticon were removed. The training set is set and it

consists of 5000 positive and 5000 negative movie reviews respectively.

## 2.4 Real Time Application Technique

As mentioned above, we use Meteor Js as our core technology to build this real-time web application. In year 2015, JavaScript is a top 8 programming language that used by most people. First, JavaScript is totally free to use. With the rise of online learning platform such as Treehouse, CodeAcademy, CodeSchool and other, the threshold of learning JavaScript is lowered down quite a lots. JavaScript is born to build a web application as it does not required to be debugged, unlike C++, Java and C#. Nowadays, there are a lots of framework or package built on top of JavaScript to even easier the process of web development. However, not all the framework are suitable to build real-time web application. There are some example have been under our consideration, such as Angular.Js, Express.Js, Backbone.Js and Meteor.Js. All these framework are built to cut down the time of building of real-time web application. Out of these frameworks, we selected Meteor.Js as our main core technology to develop our dashboard application.



- Meteor is a complete open source platform and it won't cost any cents for using it.

- Meteor takes full-stack approach that can run both on server and client and work holistically together. For normal development, we might need few kind of programming languages to complete one web application. We need back-end programming language such as Ruby, C#, PHP to run over the framework to build the web application. However, with Meteor Js, can let us build the entire system with only one language.

- Meteor has world-wide community to develop packages that ready to install and use. For simple function like user login and logout can be done by just one

single command to install the particular packages to your system. Meteor has itself package platform named Atmosphere Js which opened to public and allowed people to build and published their package.

- Meteor has strong support by community and it is constantly updated to maintain its stability. Meteor has 32k of Github stars and 15k of commits.

- Meteor support MongoDB database which can handle large amount of complicated and unstructured data.

- In tern of performance, Meteor has an outstanding result when comparing with the other frameworks discussed above. After testing, Meteor has lowest initial script loading times, second lower of rendering times and lowest in re-rendering script times on changes. Data update on screen will be as fast as blinking the eyes.

Not only this, Meteor is also deemed as the best framework for developing a full-system of web application. It is believed to be the next level of web development. Mattias Petter Johansson (2016), a software engineer in Spotify, also supporting this idea on Quora when comparing Meteor Js to Express Js, Backbone Js and Angular Js. Mattias also claimed that Backbone Js and Angular Js have the similarity which both of them are running on client side. Both of these technology will not put any focus on backend algorithm, especially the communication and transportation of data between client and server. Backbone Js is a lightweight framework and it only does of it's capabilities. However, Angular Js is the creation of Google, one of the best technology company in the world. As for sure, Angular Js is sound much convincing as it is always backed by Google. Conversely, Express Js is a created based Node.js by adding a formalized structure for building the backend for web applications. Express Js can be said that is a complete opposite technology than Backbone Js and Angular Js. Express Js come to a pretty good choice to build a REST API. In the end of the article, Mattias (2016) also mentioned why people

should use Meteor JS over Backbone Js, Angular Js and Express Js to build a full-stack web application. Although, Meteor is young and new framework comparing to others, but it is well-funded and production-ready platform. The report from Tech-Crunch, written by Kyle Russell (2015) clarified that Meteor have the total raise of $20 Millions for its stack of frameworks and tools for developing web and mobile applications in JavaScript and the raise is fully funded by Andreessen Horowitz and Trinity Ventures in series B funding. This statement also proved by Owler (2015). As Meteor is one of the few frameworks that takes full-stack approach, it is highly suggested for building a real-time web application. The application will be able to run on the server (in node on the server) and the client (in JavaScript engine on the browser) and they work perfectly. Meteor Js comes bundled with MongoDB, added by Mattias.

The explanation made by Mattias Petter Johansson have gained the total of 932 up-vote by other users so far. Under the statement of Mattias, Leo Harsha El Niño, a Meteor Developer in Geek also help in explaining the weakness of Angular Js and Backbone Js which are not suitable on building the full stack application. Leo mentioned that although Angular managed to make the development easier and faster, but the scope will only limited to client-side usage. One of the strongest feature of Angular Js is able to use re-use components since UI behavior is wrapped up in small directives. Unlike the Angular Js, Meteor is claimed to be the future of web, by Leo (2016). Leo also recommended Meteor is the best choices for startup who want to work with both front end and backend websites.



The technology and framework mentioned not only work well when standalone. However, they are all compatible with Meteor to improve the performance. The finding by Shawn McKay (July, 2015) also showed that Meteor is increasing becoming a less opinionated framework. There are more and more front-end integrations to boost up the performance of Meteor, such as React, Angular-Meteor and Angular 2. The performance is examined by 3 key measurements which are Initial Script Loading Times, Rendering Script Times and Re-rendering Script

Time on Changes.

## 2.5 Project Methodology

Selecting a project is essential as choosing a good ingredient for food. Project Management act like a map, with the step-by-step instruction to guide us to finish the the project successfully. There are so many different kind of project methodology out there, choosing a right methodology is something cannot be overlooked. After the some research was made, Waterfall methodology is finally selected.

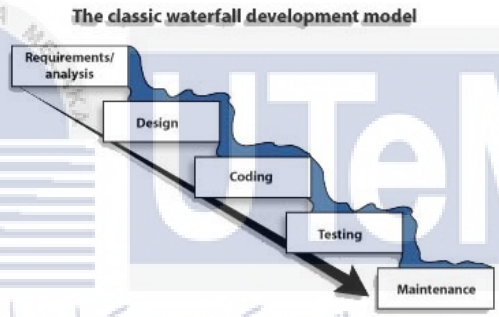




**Figure 2.0 : Classic waterfall development model**

Waterfall model can be stated is the earliest SDLC approach that was used for software development. It is also referred to a linear sequential life cycle model. The model itself can divided into many phase which each phase has it own functionality and purposes on different part on development. Meanwhile, it also has the rule of any phase can only be begin if the previous phase is completed. In waterfall model, each phase will never overlap with other phases. As advantage, waterfall model allows for increased control throughout each phase. However, if changes is made, waterfall model showed its inflexibility .

The sequential phase in waterfall model are :

1. **Requirement / Analysis** : In this phase, all possible requirements of the system are collected and written down in the requirement specification doc. Out of the 5 phases, the first phases is extremely essential to this project. The bad preparing and understanding of requirement can easily lead to failure of project. Therefore as this stage, we not only doing literature review to support our point of views. In the same time, the detail scheduling of the milestone and timeframe of the project are also planned. The documentation and instruction of using YouTube API is also studied at this stage. The model is also modified and transformed into the project flow chart.

2. **Software Design** : Before this phase started, make sure that the requirement analyze work was fully completed. System Design help in specifying hard and system requirements. This phase is not only limited to interface design, but the design of system architecture and the work flow also can be included. At this stage, we designed the project interface in the mock-up. Beside this, we also designed the system architecture for accessing YouTube API to extract the data needed into our database.



3. **Coding** : Following the algorithm and architecture that have been designed in the previous phase, the system is first developed. The project started with front-end development which turn the design into the code that needed. Functionality of system will later start with the development on back-end. Starting with the function of crawling data from YouTube, and then integrate with interface to display the data retrieved on the front-end dashboard. Added the functionality of prediction model will be the last part of development. It is needed to ensure that data is enough before the prediction model is run.

4. **Testing** : Post integration the entire system is tested for any faults and failures. System needed to be test so that it is flawless to be presented. Tons of debugging and error resolving will be done on this phase. Besides this, we also need to test the algorithm accuracy. The prediction model is tuned until

the expected result is displayed.

5. **Maintenance** : There are some unpredictable issue might turn up at the client environment. Where this phase is to fix those patches are released and also enhance the product in term of performance. on the project, we will keep update the version so that our web application is runnable on every environment, regardless desktop or mobile.

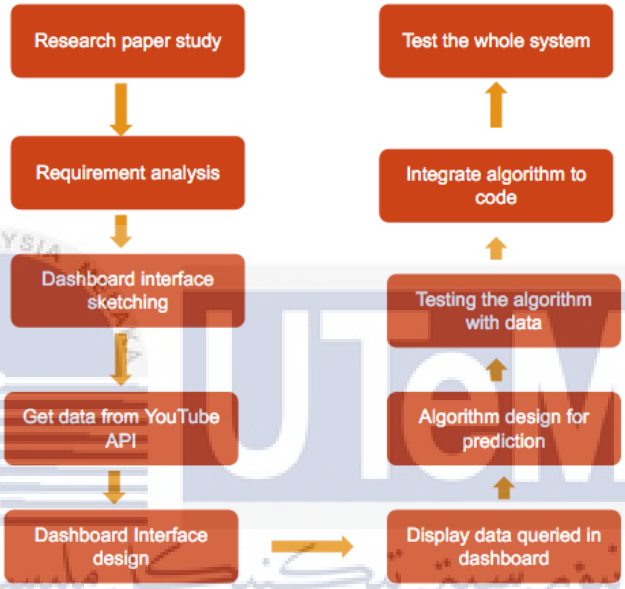




**Figure 2.1 : System development model on top of waterfall development model**

To explain the statement above, the waterfall model is modified to become the system work flow. The system will start with studying research paper which is for requirement analysis. This phrase is belong to first phrase in waterfall model. The activity followed by sketching dashboard interface in mockup and the system architecture design which is extracting the data from YouTube API. After all the designing works are completed, process will end the second phrase and proceed to third phase where the full development start. The development included front-end and back-end development, integrate two platforms to display the data on dashboard as well as develop a suitable prediction model to forecast the sale of mobile devices.

Last phase will be testing stage where system testing and algorithm testing will be carried out at this stage.

## 2.6 Project Requirements

### 2.6.1 Software Requirement

All of the softwares that used to develop the system is free and open source.

1. Atom Text Editor

   Atom is a free and open-source text and source code editor for OS X, Linux, and Windows with support for plug-ins written in Node.js, and embedded Git Control, developed by GitHub. Atom is a desktop application built using web technologies. Atom is hackable as it can allow user to customize to do anything but also user productively without changing a core configuration setting. Atom come with the built-in package manager, which mean that atom can allow user to search and install new packages created by community. Atom support smart autocompletion which can help to faster development speed. Atom comes pre-installed theme with simple yet beautiful design. Integration between Atom and version control is perfect well. High performance in operation and reliable make the starting process fast and smooth.

2. Version Control Git

   Version Control is the system that allows versioning of a project and Git is one kind of version control system in the market. Git is free and open source since it originally developed in 2005 by Linus Torvalds, the creator of Linux.

As Git is open to community, by the time goes, it is very mature now.

3. Github

   Github is a web-based Git repository hosting service. To be simple, Github is like a Google drive or dropbox for git. Every version made on the local machine will be uploaded to cloud through Github. The source code of project will remain unlovable as user can always download the latest version of code from the cloud. Students70 will given chance to apply for Github Student Developer pack which can allow students to upgrade their account to premium account. With premium account, student can have unlimited private repository which is hidden to public

4. Terminal

   Terminal is the terminal emulator included in the OS X operating system by Apple. What terminal emulator does is providing a command line interface to the operating system when used in conjunction with a Unix shell. Installation of Meteor Js on machine required terminal to be done. Command line allow user to start a new Meteor project, start and off Meteor server, update the project version, install Meteor Package to the project and many more. In this project, command line also use to output the information of data for testing purpose.

5. Safari Web Browser

   Web Browser is used to display the output of the web application. Safari is pre-install in OS X Operating System

### 2.6.2 Hardware Requirement

As this system is a web application, only one hardware require which is my working machine - Macbook Pro 15 ’inch 2012. Macbook Pro come with latest El-Capitan OS which is Unix-Based and it is designed to be as easy to use as it is beautiful to look at. Web developing can be fast and efficiency to be done running on Macbook.

## 2.7 Project Schedule and Milestones

### 2.7.1 Project Workflow

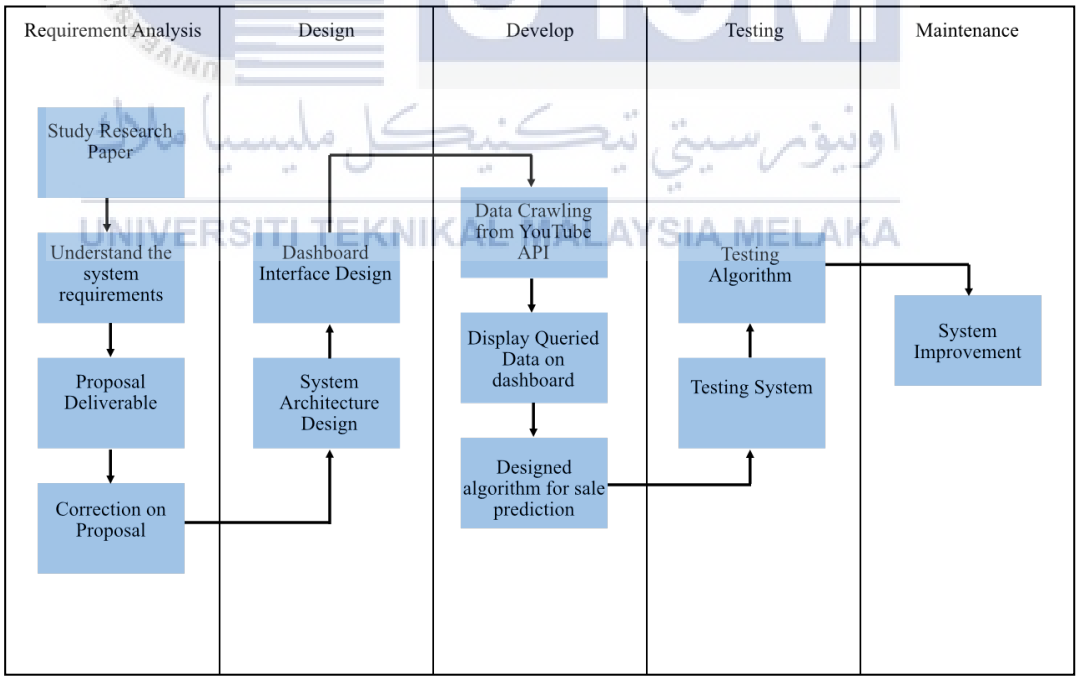


**Figure 2.2 : System Workflow Diagram**

### 2.7.2 Gantt Chart of Project Activities

**Table 2.0 : Development Gantt Chart**

| Task Name | Sub Task Name | w 1 | w 2 | w 3 | w 4 | w 5 | w 6 | w 7 | w 8 | w 9 | w 10 | w 11 | w 12 | w 13 | w 14 | w 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Requirement Analysis** | Study Research Paper | ■ | | | | | | | | | | | | | | |
| | Understand the system requirement | | ■ | | | | | | | | | | | | | |
| | Proposal Deliverable | | | ■ | | | | | | | | | | | | |
| | Correction on proposal | | | | ■ | | | | | | | | | | | |
| **Design** | Dashboard interface design | | | | | ■ | | | | | | | | | | |
| | System Architecture design | | | | | | | ■ | ■ | | | | | | | |
| **Develop** | Data Crawling from YouTube API | | | | | | ■ | ■ | | | | | | | | |
| | Display queried data on dashboard | | | | | | | | ■ | ■ | | | | | | |
| | Algorithm for sale prediction Designed | | | | | | | | | | ■ | | | | | |
| **Testing** | Testing algorithm | | | | | | | | | | | ■ | ■ | | | |
| | System testing | | | | | | | | | | | | | ■ | ■ | |
| **Maintenance** | System improving | | | | | | | | | | | | | | ■ | ■ |

**Table 2.1 : Academic Gantt Chart**

| Task Name | w 1 | w 2 | w 3 | w 4 | w 5 | w 6 | w 7 | w 8 | w 9 | w 10 | w 11 | w 12 | w 13 | w 14 | w 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Proposal Deliverable | | | | | | | | | | | | | | | |
| Enhancement on proposal | | | | | | | | | | | | | | | |
| Chapter 1 Preparation | | | | | | | | | | | | | | | |
| Chapter 1 Deliverable | | | | | | | | | | | | | | | |
| Chapter 2 and Chapter 3 Preparation | | | | | | | | | | | | | | | |
| Chapter 2 Deliverable | | | | | | | | | | | | | | | |
| Demonstration of Chapter 3 and Chapter 4 | | | | | | | | | | | | | | | |
| Preparation of Chapter 4 | | | | | | | | | | | | | | | |
| Delivery of Chapter 4 | | | | | | | | | | | | | | | |
| Deliverable and Presentation of PK2 | | | | | | | | | | | | | | | |
| Demonstrations | | | | | | | | | | | | | | | |
| Demonstrations and PSM report | | | | | | | | | | | | | | | |
| Deliverable of PSM Report | | | | | | | | | | | | | | | |
| Presentation | | | | | | | | | | | | | | | |
| Correction on Draft Report | | | | | | | | | | | | | | | |



### 2.7.3 Milestones and Date

**Table 2.2 : Project Milestones for PSM 1**

| Project Activity PSM 1 | Duration(days) | Date Completed |
|---|---|---|
| Discussion of title and proposal | 14 | 1/1/2016 - 15/1/2016 |
| Writing on Chapter 1 - Introduction | 8 | 22/2/2016 - 1/3/2016 |
| Writing on Chapter 2 - Literature Review | 10 | 2/3/2016 - 14/3/2016 |
| Writing on Chapter 3 - Methodology | 8 | 15/3/2016 - 23/3/2016 |
| Analysis | 10 | 24/3/2016 - 3/4/2016 |

| Project Activity PSM 1 | Duration(days) | Date Completed |
|---|---|---|
| Design | 5 | 2/4/2016 - 8/4/2016 |
| Mid-Semester Break | | 9/4/2016 - 17/1/2016 |
| Implementation | 30 | 18/4/2016 - 18/5/2016 |
| Testing | 5 | 19/5/2016 - 24/5/2016 |
| Presentation | 4 | 25/5/2016 - 15/6/2016 |
| Writing of PSM 1 Report | 10 | 16/6/2016 - 26/6/2016 |

**Table 2.3 Project Milestones for PSM 2**

| Project Activity PSM 2 | Duration(days) | Date Completed |
|---|---|---|
| Testing and writing Final Report | 31 | 1/7/-2016 - 31/7/2016 |
| Presentation | 4 | 1/8/2016 - 5/8/2016 |
| Submission of Final Report | 10 | 6/8/2016 - 30/8/2016 |

## 2.8 Conclusion

As the conclusion of chapter 2, deep and hard consideration is made before every decision. As the result, Meteor Js has been selected be the main core technology for developing the system. As for the development approach, waterfall model is chosen so that the system development is undergo under control. Furthermore, all the requirements including hardware and software requirements have stated as clear as possible. A detailed project schedule and milestone is also set so that the progress can be verified from time to time by referring to the activity and its respective output of every stage. Everything is set, let's get the hand dirty. On next chapter, the analysis about this project will be discussed.

# CHAPTER III

# ANALYSIS

## 3.1 Introduction

This chapter explains the analysis done on the proposed system. The system scenario and situation will be demonstrated in the form of flow diagram as well as activity diagram. The total flow of the system will also be visualized in the flow chart to give clearer understanding. Last but not least, the requirement analysis will also be carried out on this chapter and can be divided into 3 part which are Data requirement, Functional requirement and non-functional requirement. This chapter ended up with conclusion.



## 3.2 Problem Analysis

The analysis of the proposed will be explained in several diagram and it will stared with the decomposition diagram. Decomposition diagram can provide a clear visual representation of all modules in the system. Besides this, the decomposition diagram shows a high level function and its lower level components which is build to

complete the high level function. In overall, decomposition diagram will able to provide a logical hierarchical decomposition of a system.

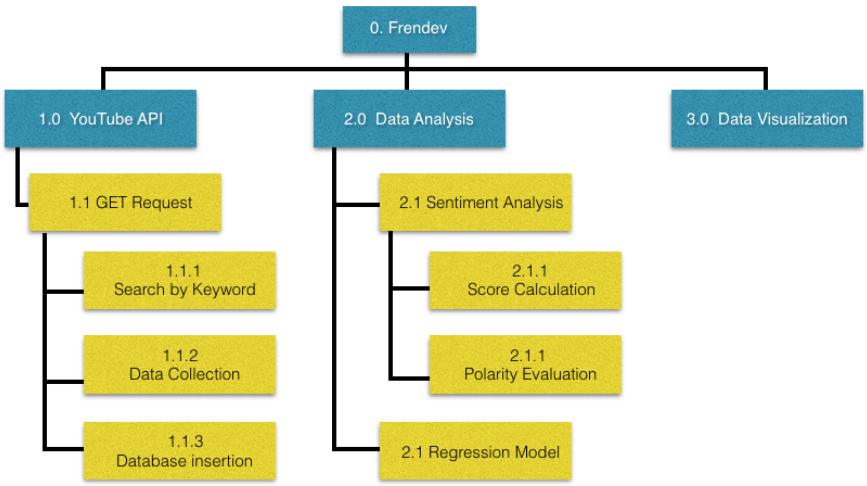


**Figure 3.1 : Project Decomposition Diagram**

Three core modules of the system is displayed and broken down into pieces of lower level sub-module. Out of these 3 sub-module, collected data from YouTube API showed the most complicated steps when comparing to the others 2 sub-module. The whole process can be broke down to many small steps. Every process for retrieving data has to be done step by step without any interference. The process chunk down into 4 core steps where each step has its own functionality to retrieve different type of data.

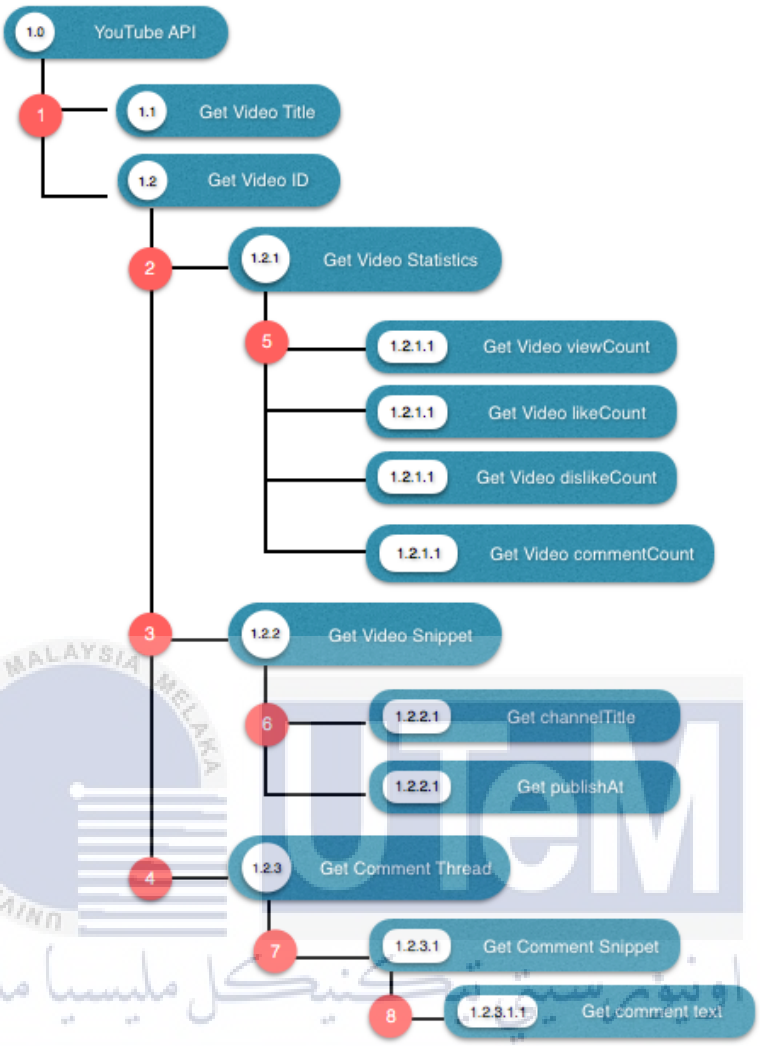




**Figure 3.2 : Module Youtube API Breakdown chart**

During step 1, 50 video titles and video id are returned and collected using YouTube search service. Those video id are next used to retrieve video statical data in step 2. Statical data collected including video's view count, like count, dislike count and comments count. Video id are highly essentials in whole process. In step 3, id also used to retrieve video snippet data which are including video's channel name and video's published date. In order to get all the comments under the video, video id is again used to collect comment thread data which will next help to get all the video's comment text for sentiment analysis. The data collected will direct insert to database right after each core step.

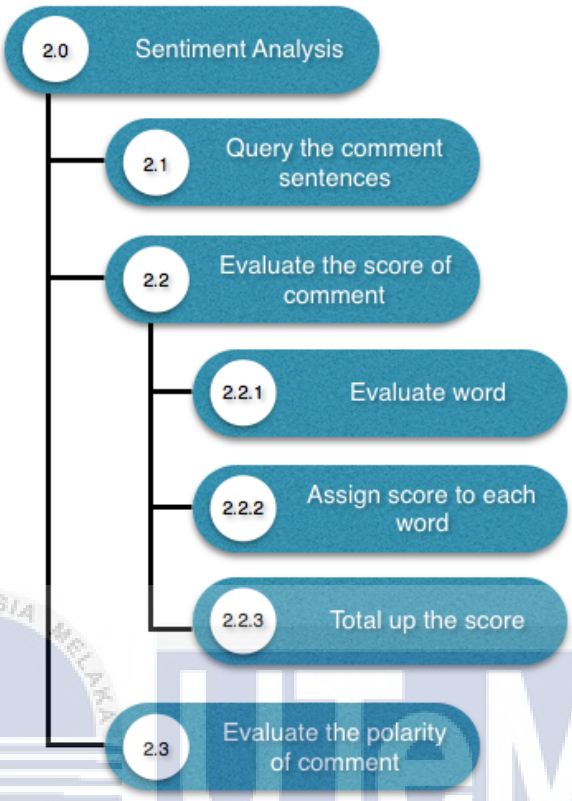


**Figure 3.3 : Module Data Analysis Breakdown chart**



After the comment is retrieved to the system by YouTube API, sentiment analysis is run. First, the system will query the comment and pass to the sentiment function. The sentiment function will evaluate the every word in the sentences and assign score to each word. By total up the score, sentiment function will return the total score of comments in variable. This variable will then evaluate to determine the polarity.

### 3.3 Requirement Analysis

#### 3.3.1 Data Requirement

All the data retrieved from YouTube API are stored in MongoDB, which is a No-SQL databases. Instead of storing data in tables and rows in relational database,

MongoDB let user to store JSON-like documents with dynamic schemas. In MongoDB, table will call as one collection to store the data. Videos will stored as one objects in a table, and comments will also stored in one objects but park under videos collection. There are total of 3 collections in the database, which are Videos, Sale and Prediction. Sale collection is used to stored all the historical data of Apple iPhone sale whereas predictions collection is used to stored the data for prediction as well as prediction result. The data stored for predictions are including average number of likeCount, dislikeCount, viewCount, and positiveRate.

### 3.3.1.2 Data Dictionary

Collection Name: Videos

| Name | Data Type | Primary Key |
|---|---|---|
| _id | String | Yes |
| title | String | |
| id | String | |
| viewCount | Int32 | |
| likeCount | Int32 | |
| dislikeCount | Int32 | |
| commentCount | Int32 | |
| channelName | String | |
| publishAt | date | |
| comment | array | |

Object Name : Comments

| Name | Data Type | Primary Key |
|---|---|---|
| Text Display | String | |
| Like | number | |
| score | Int32 | |
| positivity | string | |

Collection Name : Sales

| Name | Data Type | Primary Key |
|---|---|---|
| _id | String | Yes |
| period | string | |
| model | string | |
| keyword | string | |
| sale | int32 | |

Collection Name : Predictions

| Name | Data Type | Primary Key |
|---|---|---|
| _id | String | Yes |
| keyword | string | |
| sale | int32 | |
| likeCount | int32 | |
| dislikeCount | int32 | |
| viewCount | int32 | |
| positiveRate | int32 | |
| predictSale | int32 | |

### 3.3.1.2 Data Model


**Figure 3.4 : Data Model of whole system**

The figure showed that the data model of the full system which contains of 3 different kinds of data collections. Each box in the diagram represent each collections which are videos collection, sale collection and prediction collection. Videos collection is used to store the information of video that retrieved from YouTube API. There is a grey box which is represent the comments objects that stored under videos collections. The result of sentiment analysis including score and positivity of the comment will stored under the comments object. It is also included the comment text and like score. For sale collections, the historical sale data of iPhones will be stored in this collections. Meanwhile, prediction collections is responsible to store the data that ready to run in predictive model as well as the predicted result.

### 3.3.2 Functional Requirement

There are many functions that are needed in this system to retrieve data from YouTube API. Therefore, the significant function is described in this section to present the process of the first data collecting until the sentiment analysis.

1. fetchFromService

   This function is used to retrieved video id and video title by using keyword search. The request is given 30000 milliseconds to response before the request timeout. The id and title collected will passed to the following function.

2. insertVideo

   This function is created to store the video id and video title to database right after they are returned from API.

3. getVideoStatistics

   With the video id storing in database, this function can used to retrieve the statical data of the video. The statical data collected including video's view count, like count, dislike count and comments count and all of them will be stored inside the database under video collection.

4. getVideoSnippet

   With the video id storing in database, this function can used to retrieve the statical data of the video. The snippet data collected including video's channel id and video's published date and both of them will be stored inside the database under video collection.

5. getChannelData

   By using the channel id, this function allow the system to get more

information about the channel, especially the channel name.

6. getVideoComment

   This function is used to get all the comments under video. Sending the request by video id to get the snippet data of comments. Each data will containing text display, likeCount and published date. These 3 data will be stored in the comment array and pushed to video collection. As each response only hold up to 50 comments, the system need capture the page token in order to get the next 50 comments and the loop will be ended when there is no page token returned

7. SentimentAnalysis

   Before the comments are stored in the video collection, sentiment analysis function is performed. After the process, the score and positivity of the comment will be updated to the comment array, and push together to the video collection.



### 3.3.3 Non-Functional Requirements

1. Performance

   Performance of data collecting process can be measured by the response time from the API. The timeout of 3000 millisecond is for single response. This can help to establish the connection. Time won't be wasted if no data are received.

2. Data Integrity

   All the data collected from YouTube API will not be modified, edited and changed. Remaining the originality of the data is one of the fundamental of

the system.

3. Manageability

   The data collected is directly showed in the data collection table. High manageability on the data.

### 3.4 Conclusion

Making a goof literature review is like showing a path to bring us to our target. Research and understand the system requirements prepare us to overcome the problem might be faced during the development process. Well planning on database with structured data dictionary and data model allowed us to understand what data is needed to be used and that can prevent wastage of usage of database. With right arranging of data, also help to speed up the data transactions time.

Next chapter, we will discussed about the design of the system, including high-level design, framework design, interface design, technical design, and more.

# CHAPTER IV

# THE PROPOSED SYSTEM DESIGN

## 4.1 Introduction

System Design is the process of defining the system architecture design, user interface design and database design. In the user interface design section, it covers navigation design, input and output design. Except from this, conceptual and logical database design also briefly introduced in this chapter. The more detailed design such as Software Design and Physical database design also included.



## 4.2 High-Level Design

In this section, system architecture and interface design will be discussed and explained with the diagram.

### 4.2.1 System Architecture for system

The system architecture used in the development of this project is three tier architecture. The system is a client-server architecture which made up of presentation tier, logic tier and also data tier. This architecture is commonly to be seen in any web application.

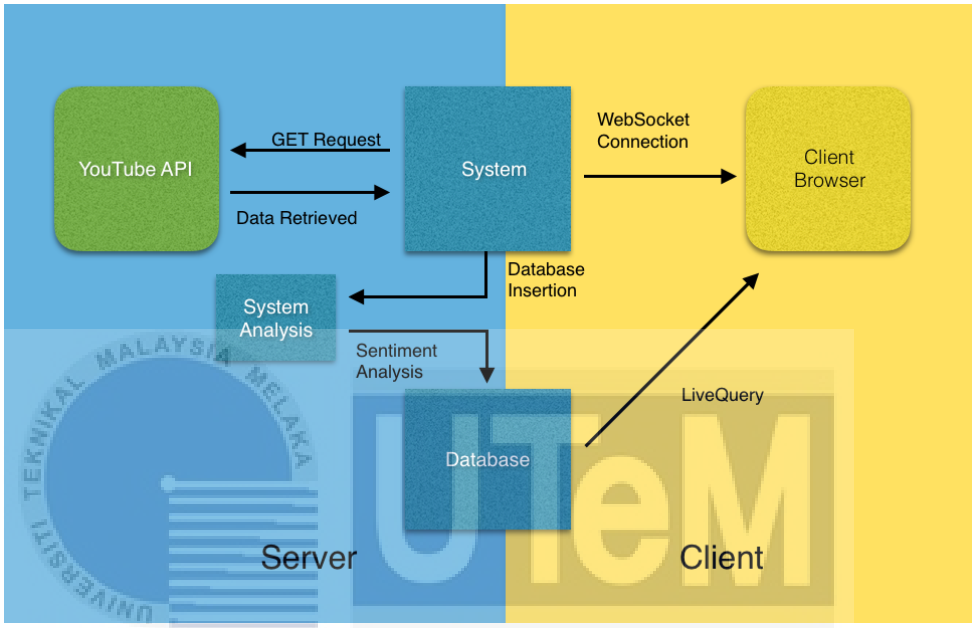


**Figure 4.1 : System Architecture Diagram**



The presentation tier is top level tier of the application which is in client side. It present the UI interface which can interact with user. Translate and visualize the text form data into the beautiful design of dashboard. On this data hand, server side involved both logic tier and data tier which is not allowed to be presented to the user. Logic tier responsible to handle all the system logic, for example query handling, error handling, routes, data analysis and more. Logic tier structured the data and send it to the presentation tier. Lastly, the data tier is located at bottom layer of the system. Basically, in this tier, data is transferred in and out from there database. The data extracted from the API will be stored in database after filtered by logic tier.

Not to mentioned, as the system is built as a dynamic web application, which

not only well on displaying and visualizing the data on the interactive dashboard, the algorithm running back-end to perform the analytic.

### 4.2.2 User Interface Design for system

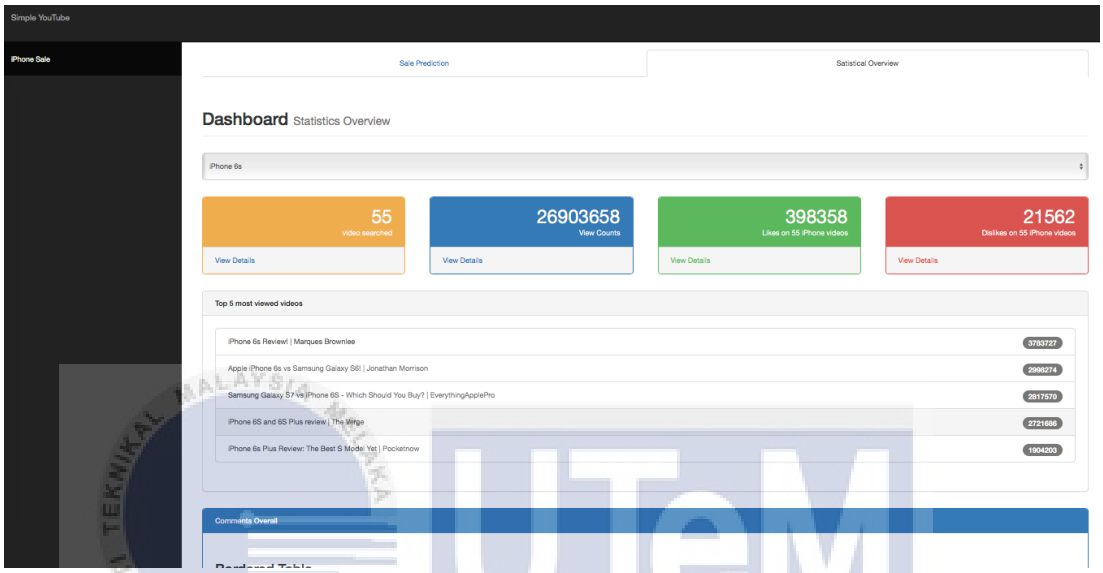


**Figure 4.2 : User Interface of main page of the system 1 (upper part)**

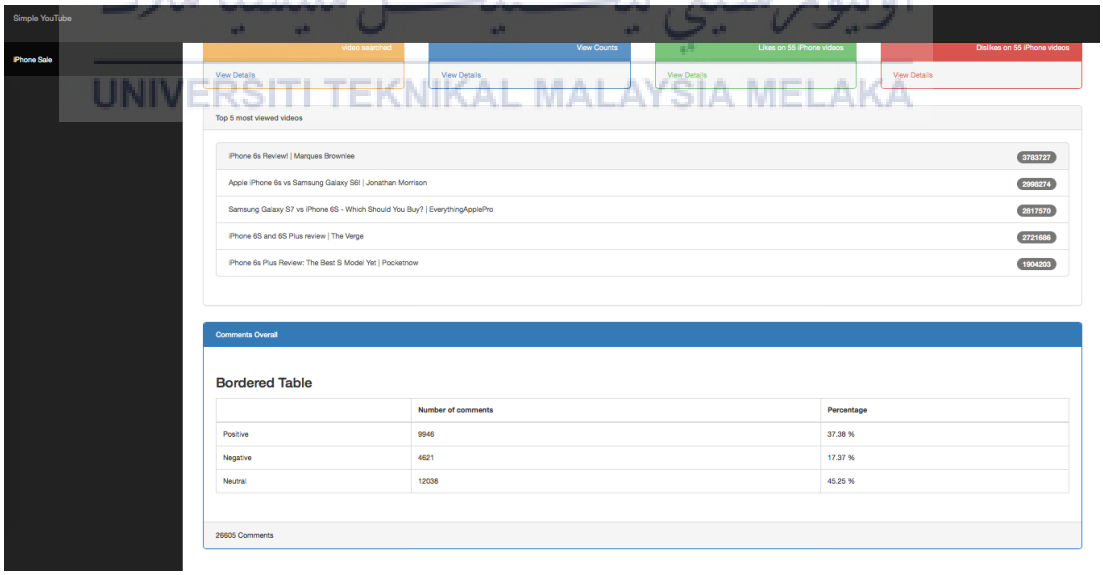


**Figure 4.3 : User Interface of main page of the system 2 (lower part)**

The user interface of the system is designed with front-end framework bootstrap and the dashboard design is integrated with SB Admin dashboard. The main user interface is easy to understand as it just visualize some basic information about the video. As we can see, there are some boxes design showing the number of videos collected, total number of like count, dislike count and views count. On the center part of main page, there is the list table sorting the channel name in descending order with the view of videos. At the lowest part of the system is where the comment data showed. The table used to show the overall comments data which already categorized into 3 group, positive, negative and neutral.

#### 4.2.2.1 Navigation Design

To keep the application easy to use, the system is design to reduce the number of the pages. The small navigation bar with the basic functionality which will only redirect the page to either statical page (main page) or data prediction page.


**Figure 4.4 : Navigation Design of the system**

By clicking the sale prediction button, the page will redirect to prediction pages. Besides this, there are also an option bar on the main page to change the video content. Users are free to change the value on the option bar, the data will change immediately without refreshing the page when the value on the bar is changed.

#### 4.2.2.2 Input Design

There are no input to insert to the system. However, there are step that need to be done when user want to collect new data.

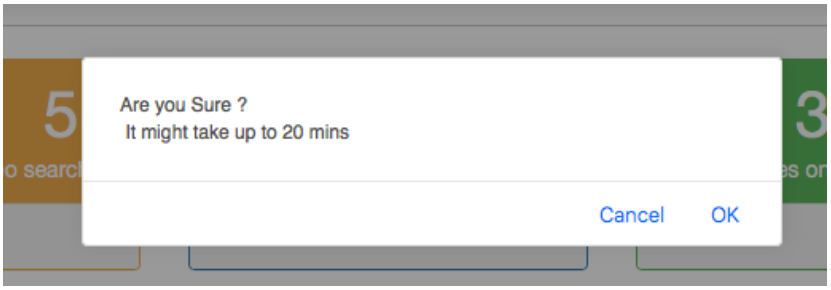


**Figure 4.5 : Alert Notification before the data collection start**

By clicking the "iPhone sale" button on the sidebar, the notification will pop out to confirm the process. The system didn't provide any input method. At this early stage of development, wrong input data might cause the collapse of the system.

#### 4.2.2.3 Technical Design



AI technique that applied in the system is sentiment analysis and the algorithms for predicting the sale is multiple regression model.

```
1. Get the video comment in text form
2. For each comments in video collection
   1. Check word by word in the comment
      1. For each word in comment
         1. Assign the score to the specific word according to the AFINN text library
      2. End Loop
   2. Sum up all the score of the comments.
   3. Assign the popularity to the comments.
3. End Loop
```

**Figure 4.6 : Pseudo-code for sentiment analysis**

Sentiment analysis can also be known as opinion mining which is perform of natural language processing and text analysis to identify and categorize opinions expressed in a piece of text. Generally, sentiment analysis aims to determine understand the writers attitude towards a particular topics and determine whether this attitude is positive, negative or neutral.

The technique used in this project is used with the Node.js module named *Sentiment* that created by *thisandagain* on Github. *Sentiment* uses the AFINN-111 wordlist to perform sentiment analysis on arbitrary blocks of input. In result, the module will return score and comparative of the sentences. Sentiment also provide the ability to append and overwrite values from AFINN by simply injecting key/ values pairs into a sentiment method.

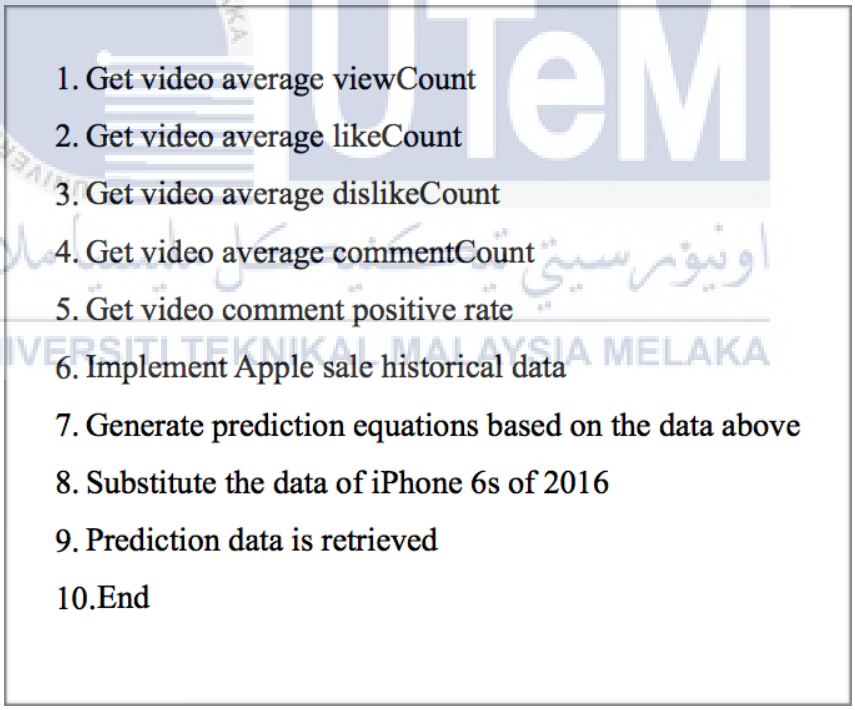




**Figure 4.7 : Pseudo-code for Sale Predictions**

Linear regression is an approach for modeling the relationship between 2 variable which is normally consist of one dependent variable and one explanatory variables or one independent variable. For example, The formula y = ax + c, is the

simples form of linear regression formula to explain the relationship between one dependent, y and one independent variable, x. Multiple regression is an extension of simple linear regression. When the number of independent variable is more than 2 variable, multiple regression is used for that particular situation.

Linear Regression is commonly used in predictive analysis. By fitting a group of data points on a single line through a scatter plot to understand the relationship between the dependent variable and one or more independent variables. In this stage, the independent variables are 5 statical information about iPhone videos which consists of average view count, average like count, average dislike count, average comment count and comment's positive rate. Meanwhile, the dependent variable will be Apple sale historical data.

**Table 4.1 : Regression model data points**

| Phones | Avg View Count | Avg Like Count | Avg Dislike Count | Avg Positive Rate | Sale |
|---|---|---|---|---|---|
| iPhone 1 | 866228.83 | 5749.17 | 1680.50 | 30.13 | 1.39 |
| iPhone 2G | 186445.20 | 585.00 | 94.60 | 12.96 | 11.63 |
| iPhone 3G | 269808.86 | 969.14 | 74.54 | 38.29 | 20.73 |
| iPhone 3Gs | 161408.76 | 633.90 | 76.92 | 30.87 | 39.99 |
| iPhone 4 | 553959.89 | 2424.93 | 175.41 | 33.59 | 72.3 |
| iPhone 4s | 437542.64 | 2070.86 | 153.88 | 30.94 | 125.04 |
| iPhone 5 | 691300.80 | 4005.20 | 514.95 | 37.38 | 150.26 |
| iPhone 5s | 390919.10 | 4440.50 | 244.08 | 33.52 | 169.19 |
| iPhone 6 | 635100.52 | 7604.42 | 409.26 | 31.91 | 231.22 |
| iPhone 6s | 489149.58 | 7242.85 | 392.04 | 37.38 | 125.97 |

After identifying the strength of the effect that the independent variables have on a dependent variable, the predictive model is built.

$$y = (0.000080075 * A) + (0.034324 * B) + (-0.15822 * C) + (0.53075 * D) - (9.7058)$$

whereas

y = iPhones sale A = Average view Count

B = Average Like Count C = Average Dislike Count

D = Average Positive Rate

With the generation of the predictive formula, the understand on how much will the dependent variable change when one or more independent variables showed a clear picture. In the next stage, the regression analysis can be used to get point estimate which is used to estimate the sale since the predictive equation is built. By substitute the iPhones 6s statical data back to equation, the answer is shown. Multiple regression help to determine the overall fit of the model and the relative contribution of each of the predictions to the total variance explained. From the predictive equations, we can understand that comments positive rate and average dislike count reflect an heavy weight on affecting the sale production. However, the view count number and the like count showed least effect on the sale production. 9.7058 is the constant value for the equations.

4.2.2.4 Output Design

The prediction data will be displayed in the line graph on the data prediction pages and also a table with all the data that needed to run the regression algorithm. Last components in the prediction page in donut chart and histogram chart to display the result of sentiment analysis.

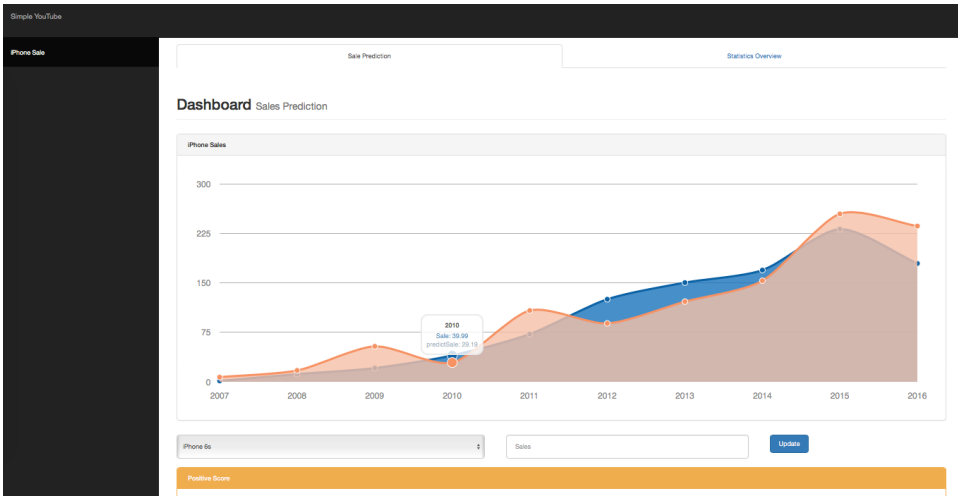


**Figure 4.6 : Prediction data on the line graph**

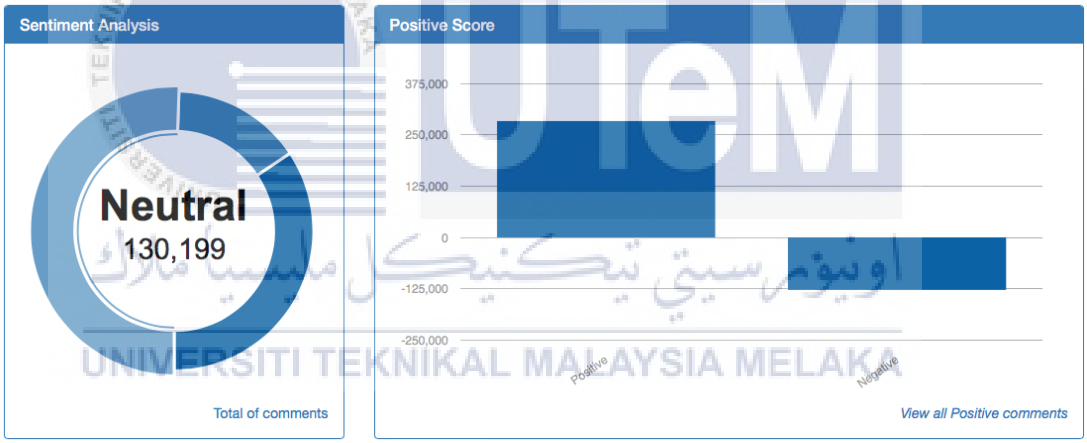


**Figure 4.7 : Result of Sentiment Analysis on the donut chart and histogram**

## 4.3. Conclusion

In this chapter, the system architecture design is explained how the system work on the back end development. Besides this, the front-end development progress also been discussing with the user interface. The overall workflow for the artificial intelligence technique are written in pseudocode to ease the process of understanding

the code. System design playing a big role to produce a success system. Next chapter, we will discussed the implementation state.

# CHAPTER V

# IMPLEMENTATION

## 5.1 Introduction

This chapter discusses about the implementation of the proposed system. It also cover the software development setup, software configuration management and implementation status of every module.



## 5.2 Software Development Environment Setup

Once Meteor is installed in the computer, the environment will automatically set up. By starting the new project, Meteor will use MongoDb as the primary database by default. In this project, Meteor will be setting up at version 1.3.1 on top of node.js v0.10.43 whereas the default database is running on MongoDB version 2.6.7 .

**Table 5.1: Meteor Server Status**

| Key | Value | Type |
|---|---|---|
| host | JiaJing-MacbookPro.local:3001 | String |
| version | 2.6.7 | String |
| process | mongod | String |
| pid | 29927 | Int64 |

**Table 5.2 : Server and Database Configuration in Development Environment**

| | |
|---|---|
| Port Number | 3001 |
| IP Address | 127.0.0.1 |
| Database Instances Name | meteor |
| Hostname | localhost |
| Database Name | null |
| Database Password | null |



## 5.3 Software Configuration Management

### 5.3.1 Configuration Environment Setup

In web application deployment, it is common to refer to three runtime environment which are Development, Staging and Production. In this project, all the deployment will only run on the first stage which is on Development. This refers to the machine where the systems is developed. Starting a server in Meteor is as easy as writing a line of command line on terminal.

The environment will automatically setup once the server is start running. Just by typing “meteor” , after changing directory to your current project file, everything is set.

```
~/Meteor/youtube3(branch:rearrange-page*) » meteor
[[[[[ ~/Meteor/youtube3 ]]]]]

=> Started proxy.
=> Started MongoDB.
=> Meteor 1.3.2.4 is available. Update this project with 'meteor update'.
=> Started your app.
```

**Figure 5.1 : Software Configuration Setup**

Terminal is an interface in which you can type and execute text based commands on Mac OS. We can think of terminal as command line on Window OS. To setup Meteor environment, terminal is an only tools needed.

1. Change Directory to the Meteor folder by

   *cd Meteor*

2. Change Directory to the system folder by

   *cd youtube3*

3. Start the meteor server by

   *meteor*



### 5.3.2 Version Control Procedure

As mentioned before, Version control allow user to have versions of a project which show the changes that were made over times. In this project, git is selected as the primary version control system that help to track every change that made. To set up the git in the project is as easy as typing one command. The system also require terminal to perform git.

1. Change Directory to the Meteor folder by

   *cd Meteor*

2. Change Directory to the system folder by

   *cd youtube3*

3. Initialize the new repository of git to the system

   *git init*

Executing the command above, will create .git subdirectory in the project root. The .git directory is a hidden folder and it contains all of the necessary metadata for the repository. All the changes on the system will be recorded by the system in the folder. Otherwise, an existing project remains unaltered, aside from the .git directory.

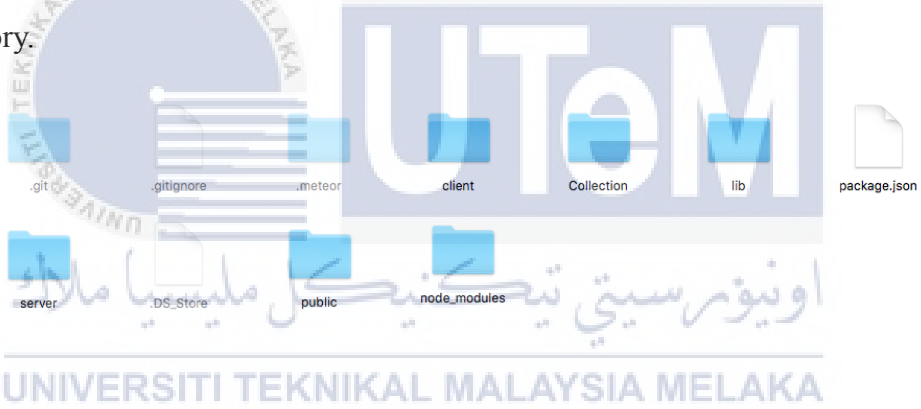




**Figure 5.2 : Overview of the system directory (.git directory is implanted)**

After setting up git in local machine, create a remote repository on github for backup purpose. This can help to recover all the data and information, even the local machine is broke down into pieces. Following steps allow us to link the local git to github.

1. Change Directory to the Meteor folder by

   *cd Meteor*

2. Change Directory to the system folder by
   *cd youtube3*

3. Create a new repository on github

4. Make a commit on the changes of the system
   *git commit -m "first-commit"*

5. Link the local git to the remote on github
   *git remote add origin* https://github.com/username/projectname.git

6. Push the local repository to github
   *git push -u origin master*

Every time there are some changes in the working directory, run the following commands to achieve the objective to track on the changes.

1. Stage all the changes for the next commit
   *git add .*



2. Commit the staged snapshot to the project history with the message
   *git commit -m "message"*

3. Push all the changes made to the github remote repository
   *git push*

By using version control, can allows us to have better qualification of code. For instances, in this project, as every time new function will be developed in this new branch and then only merged to master branch. With this method, the master branch will always have ready-function of product which has no worry on failure of system causing by the bug of other module.

## 5.4 Implementation Status

The implementation status of 3 core modules together with the sub-module will be listed in the table below. The details of each modules like description, completion data and duration will also included in the table

**Table 5.3 : Implementation Status of module YouTube API**

| Name | Task Description | Duration |
|---|---|---|
| HTTP Search Request | Search the collections of video based on the type of phones | 9 March 2016 |
| Get Video Title and Video Id | Retrieve and store video id and video title to video collections. | 9 March 2016 |
| Stop Search Request | Get only 50 response on every search request | 9 March 2016 |
| HTTP Video:LIst Request | Search Video statistics data by video_id | 10 - 12 March 2016 |
| Get Video Statistics Data | Retrieve and update video with statistical data | 10 - 12 March 2016 |
| HTTP Video:List Request II | Search Video Snippet data by video_id | 10 - 12 March 2016 |
| Get Video Snippet Data | Retrieve and update video with snippet data | 10 - 12 March 2016 |
| HTTP Comments:List Request | Return a list of comments snippet data that match video_id | 15 - 20 March 2016 |
| Get Comment Snippet data | Retrieve and store the snippet data in the array and push to video collection | 15 - 20 March 2016 |
| HTTP Comments:List Request II | Return comment Token for more comments | 15 - 20 March 2016 |
| HTTP Comments:List Request III | use nextPageToken to send new request to get more comments snippet data that match video_id | 15 - 20 March 2016 |
| Stop HTTP Comment Request | Stop the request when no new nextPage Token is return | 15 - 20 March 2016 |

**Table 5.4 : Implementation Status of module Data Analysis**

| Name | Task Description | |
|---|---|---|
| Score of comments Evaluation | Evaluate each word in the comments with score, and total up the scores for whole sentences | 5 - 9 April 2016 |
| Polarity of comment Evaluation | Assign the polarity as negative, positive or neutral to the comment | 5 - 9 April 2016 |
| Regression Model Generetion | By using the data collected which are viewCount, LikeCount, DislikeCount, NumberOfComment, CommentPolarity to train a regression model | 12 April 2016 |
| Sale Data Prediction | By substituting the data of iPhone 6 to the regression equation to predict the new sales amount of Apple | 12 April 2016 |

**Table 5.5 : Implementation Status of module Data Visualization**

| Test Data | Data Description | Expected Result |
|---|---|---|
| Data Visualization on Panel | Display total videos, video's viewCount, video's likeCount, video's dislikeCount on panel | 26 - 29 April 2016 |
| Data Visualization on Table | Display video comment number and percentage on data table | 26 - 29 April 2016 |
| Data Visualization on Line Chart | Display Apple sale history from 2007 to 2015 and sale prediction data on morris line chart | 26 - 29 April 2016 |
| Data Visualization on Table II | Display overall video data which are average viewCount, average likeCount, average dislikeCount, average positiveRate, and sale data for each iPhone from 2007 to 2016. | 3 May 2016 |
| Data Visualization on Donut Chart | Display result of sentiment analysis on Donut Chart which will be divided to 3 division - Positive, Negative and Neutral for all Apple iPhone | 3 May 2016 |
| Data Visualization on Bar Chat | Display percentage of negative comment and positive comment as well as the total number of those comments. | 3 May 2016 |

## 5.5 Conclusion

This chapter covers all the software environment setup and software configuration setup. Thanks to the simplicity and easy use of Meteor, the both setup doesn't take much time to be done. In this chapter, the version control also been discussed. All the implementation status of the modules also stated clearly in this chapter.

Next chapter, the testing process will be discussed.

# CHAPTER VI

# TESTING

## 6.1 Introduction

In this chapter, testing progress will be discussed. Making a test on the software is not only help to determine the defective or bug found in the system. Meanwhile writing can allow us to learn about the reliability of the software and ensure it work as what it planned to be. Test plan is like a rule book to guide out thinking. There are many important aspects that to be documented in this chapter, such as test estimation, test scope, test strategy and test result.



The functional testing of the system involved unit testing, API testing and integration testing. On the other hand on user experience, several activities like user testing and user interview are undergo to ensure the usability and accessibility of the system are friendly to the user.

The testing result and analysis of the result will also included in this chapter and the chapter will be ended up with conclusion of the testing progress.

## 6.2 Test Plan

### 6.2.1 Test Organization

The whole functional process will be tested by Loh Jia Jing. 2 FTMK students which are Syafiq Hisham and Aiman Zakwan are invited to run the test on user interface and user experience testing.

### 6.2.2 Test Environment

The project will be tested on the environment with Google Chrome and Firefox. and later should be available to each tester on the Macbook Pro owned by Loh Jia Jing. Firmware and Environment will be set up as stated in development state. Meteor will be setting up at version 1.3.1 on top of node.js v0.10.43 whereas the default database is running on MongoDB version 2.6.7.

As for the user interface and experience testing will be conducted as the same Macbook, and the interview result will be recorded by using Google Form.



### 6.2.3 Test Schedule

**Table 6.1 : Table of Test Schedule**

| No. | Task Name | Duration |
|---|---|---|
| 1 | HTTP Search Request | 7 July 2016 |
| 2 | Get Video Title and Video Id | 7 July 2016 |
| 3 | Stop Search Request | 7 July 2016 |
| 4 | HTTP Video:LIst Request | 8 July 2016 |
| 5 | Get Video Statistics Data | 8 July 2016 |
| 6 | HTTP Video:List Request II | 8 July 2016 |

| No. | Task Name | Duration |
|---|---|---|
| 7 | Get Video Snippet Data | 9 - 10 July 2016 |
| 8 | HTTP Comments:List Request | 9 - 10 July 2016 |
| 9 | Get Comment Snippet data | 9 - 10 July 2016 |
| 10 | HTTP Comments:List Request II | 9 - 10 July 2016 |
| 11 | HTTP Comments:List Request III | 9 - 10 July 2016 |
| 12 | Stop HTTP Comment Request | 9 - 10 July 2016 |
| 13 | Score of comments Evaluation | 17 July 2016 |
| 14 | Polarity of comment Evaluation | 17 July 2016 |
| 15 | Regression Model Generetion | 17 July 2016 |
| 16 | Sale Data Prediction | 18 July 2016 |
| 17 | Data Visualization on Panel | 22 July 2016 |
| 18 | Data Visualization on Table | 22 July 2016 |
| 19 | Data Visualization on Line Chart | 22 July 2016 |
| 20 | Data Visualization on Table II | 23 July 2016 |
| 21 | Data Visualization on Donut Chart | 23 July 2016 |
| 22 | Data Visualization on Bar Chat | 23 July 2016 |



### 6.3 Test Strategy

A well project management approach can bring big a huge impact on testing result. Not only result, a good project management approach also ensure that the testing result can be deliverable in time and estimating the resources. As there are only one tester in this project, top-down approach in the testing process when comparing with bottom-up. Although bottom up approach has more flexibility and agility, it require a lots of contribution from all team members to deliver a good and creative solutions. As there are only one tester for this project, top-down approach will be the best option.

Besides this, white-box testing will be applied to all the functional testing, whereas the black-box testing will be used while running user experience testing. White box test can ensure that all function can work flawlessly. As black-box tests are done from a user's point of view, these can be strongly help in exposing differences in the specifications. Black-box is the avoidance of developer-bias. Black-box test is conducted by a body independent from the developer of the system.

## 6.4 Test Implementation

Module : YouTube API

Made by: Loh Jia Jing

Date : 7 July 2016

| Test Case | Test Description | Expected Result |
|---|---|---|
| HTTP Search Request | Search the collections of video based on the type of phones | Return collection of search results that match the query parameters |
| Get Video Title and Video Id | Retrieve and store video id and video title to video collections. | Create 2 new objects in video collection and store with data collected |
| Stop Search Request | Get only 50 response on every search request | The request will stop after 50 videos are received |
| HTTP Video:LIst Request | Search Video statistics data by video_id | Return video viewCount, LikeCount, DislikeCount and commentCount |
| Get Video Statistics Data | Retrieve and update video with statistical data | Create 4 new objects and update with data collected |
| HTTP Video:List Request II | Search Video Snippet data by video_id | Return video channel_id and video published date |

| | | |
|---|---|---|
| Get Video Snippet Data | Retrieve and update video with snippet data | Create 2 new object and update with data collected |
| HTTP Comments:List Request | Return a list of comments snippet data that match video_id | Return comment_id, comment_textDisplay, comment_likeCount and comment publish date. |
| Get Comment Snippet data | Retrieve and store the snippet data in the array and push to video collection | |
| HTTP Comments:List Request II | Return comment Token for more comments | Return comment nextPageToken |
| HTTP Comments:List Request III | use nextPageToken to send new request to get more comments snippet data that match video_id | Return comment_id, comment_textDisplay, comment_likeCount and comment publish date. |
| Stop HTTP Comment Request | Stop the request when no new nextPage Token is return | Process to next video |

Module : Data Analysis

Made by: Loh Jia Jing

Date : 17 July 2016



| Test Data | Data Description | Expected Result |
|---|---|---|
| Score of comments Evaluation | Evaluate each word in the comments with score, and total up the scores for whole sentences | comment scores is calculated |
| Polarity of comment Evaluation | Assign the polarity as negative, positive or neutral to the comment | polarity is assigned |
| Regression Model Generetion | By using the data collected which are viewCount, LikeCount, DislikeCount, NumberOfComment, CommentPolarity to train a regression model | Regression Equation is generated |
| Sale Data Prediction | By substituting the data of iPhone 6 to the regression equation to predict the new sales amount of Apple | return the predicted sale |

Module : Data Visualization

Made by: Loh Jia Jing

Date : 22 July 2016

| Test Data | Data Description | Expected Result |
|---|---|---|
| Data Visualization on Panel | Display total videos, video's viewCount, video's likeCount, video's dislikeCount on panel | The 4 panels is shown in main page |
| Data Visualization on Table | Display video comment number and percentage on data table | The table is displayed in main page |
| Data Visualization on Line Chart | Display Apple sale history from 2007 to 2015 and sale prediction data on morris line chart | The Line Chart is displayed in prediction page |
| Data Visualization on Table II | Display overall video data which are average viewCount, average likeCount, average dislikeCount, average positiveRate, and sale data for each iPhone from 2007 to 2016. | The table is displayed in prediction page |
| Data Visualization on Donut Chart | Display result of sentiment analysis on Donut Chart which will be divided to 3 division - Positive, Negative and Neutral for all Apple iPhone | The donut chart is displayed in prediction page |
| Data Visualization on Bar Chat | Display percentage of negative comment and positive comment as well as the total number of those comments. | The bar chart is displayed in prediction page |



The usability test conducted during the user assessment of the project consist of an introduction, 3 tasks and a post-test questionnaire. Participants are instructed to share their thought with the text organizer throughout the test.

Session were performed on an individual participant with session lasting approximately 30 minutes. The whole session will be recorded down to ease the process of analyzing the result in the end of the session. Basing on the test strategy mentioned, 3 set of tasks have formed and assigned to users during the test.

Module : Usability Test

Made by: Loh Jia Jing

Date : 22 November 2016

| Test Task | Task Description | Expected Result |
|---|---|---|
| New Data Collection | Participant is required to collect new data of iPhone 6s from YouTube and all the data will be updated and stored in the Mongo database | Participant click on the “iPhone 6s ” button on sidebar to retrieve new data. |
| Solution of Prediction Data | Participant need to find out the sale prediction data for iPhone 6s 2016 by accessing to sale prediction page and study the prediction line graph | Participant is expected to click on the “Sale Prediction” button on top menu to view the prediction chart. |
| Comprehension on Statical Data Overview | Participant need to able to perform and tell every statical data of all iPhones from 2007 to 2016. The statical data are including viewCount, likeCount, dislikeCount, Top 5 viewed video and comment details. | Participant is expected to click on the “Statical Data Overview” button on top menu to view the prediction chart. By changing the title in the option bar to view different type of data. |



## 6.5 Test Result

**Table 6.2 : Table of Test Result**

| Test Case | Testing Date | Real result (OK / FAIL) | Data Type |
|---|---|---|---|
| **YouTube API** | | | |
| HTTP Search Request | | OK | MOCK |
| Get Video Title and Video Id | | OK | REAL |
| Stop Search Request | | OK | - |
| HTTP Video:LIst Request | | OK | REAL |
| Get Video Statistics Data | | OK | REAL |

| | | | |
|---|---|---|---|
| HTTP Video:List Request II | | OK | REAL |
| Get Video Snippet Data | | OK | REAL |
| HTTP Comments:List Request | | OK | REAL |
| Get Comment Snippet data | | OK | REAL |
| HTTP Comments:List Request II | | OK | REAL |
| HTTP Comments:List Request III | | OK | REAL |
| Stop HTTP Comment Request | | OK | - |
| **Data Analysis** | | | |
| Score of comments Evaluation | | OK | REAL |
| Polarity of comment Evaluation | | OK | REAL |
| Regression Model Generetion | | OK | REAL |
| Sale Data Prediction | | OK | REAL |
| **Data Visualization** | | | |
| Data Visualization on Panel | | OK | |
| Data Visualization on Table | | OK | REAL |
| Data Visualization on Line Chart | | OK | REAL |
| Data Visualization on Table II | | OK | REAL |
| Data Visualization on Donut Chart | | OK | REAL |
| Data Visualization on Bar Chat | | OK | REAL |

**Table 6.3 : Test Result of Regression model**

| Phones | Sale, A | Predicted Sale, B | Sale Mean | B - A | B - Mean |
|---|---|---|---|---|---|
| iPhone 1 | 1.39 | 7.09 | 91.31 | 5.7 | -84.22 |
| iPhone 2G | 11.63 | 17.21 | 91.31 | 5.58 | -74.1 |
| iPhone 3G | 20.73 | 53.69 | 91.31 | 32.96 | -37.62 |
| iPhone 3Gs | 39.99 | 29.19 | 91.31 | -10.8 | -62.12 |
| iPhone 4 | 72.3 | 107.96 | 91.31 | 35.66 | 16.65 |
| iPhone 4s | 125.04 | 88.49 | 91.31 | -36.55 | -2.82 |
| iPhone 5 | 150.26 | 121.49 | 91.31 | -28.77 | 30.18 |
| iPhone 5s | 169.19 | 153.19 | 91.31 | -16 | 61.88 |
| iPhone 6 | 231.22 | 254.35 | 91.31 | 23.13 | 163.04 |
| iPhone 6s | 179.23 | 235.88 | | | |



Y mean = 91.31

Square Error $_{\text{regression line}}$ = 5488.85

Square Error $_{\text{y mean}}$ = 49465.16

R-Squared = 1 - 5488.85 / 49465.16

= 0.889

R-Squared is a one type of regression validation method of analyzing the fitness of data on the regression line. It is also know as the coefficient of determination for multiple regression to measure how close of the data are to the fitted regression line. After the calculating, the answer of R-Squared is 0.889 which is absolutely showing a great fitness. In general, the higher the R-squared, the better the model fits your data. The number of sale of iPhone 6s is predicted to be 235

millions unit in 2016. However, in this stage, this is impossible to test out the accuracy of this amount is correct as Apple itself hasn't released their final quarter report of 2016. For the first 3 quarter, Apple has already sold 179.23 millions unit of iPhone so far. In order to reach the estimated sale, the up-coming financial quarter report need to show at least 50 millions unit of iPhone in these last 3 months. Research show that there are a lots of Wall Street analyst are predicting a decline in iPhones sales. Their estimates differ only in the scale of the decline (flat to -20%) and the timing of the decline. By comparing the system predictive sale with their estimation result, can give us rough image on how accurate is the prediction result.

**Table 6.4 : Analyst Prediction Result**

| Analyst Company | Prediction Sale (million units) |
|---|---|
| Morgan Stanley | 218 |
| Union Bank of Switzerland | 220 |
| Raymond James | 229 |
| Pacific Crest | 218 |

There are small differences between what Wall Street analyst prediction and the system which are only 7.79% difference. Both analyst system which using different kind of sale forecasting methods showing a same result. There are about 10% of decline in iPhone sale in 2016 compared to 2015. Therefore, we can concluded that 2016 is looking like it will be a bumpy ride for Apple.



## 6.6 Conclusion

In this chapter, we performed all kind of testing on the system which including white box and black box texting to test all the functionality of the system. Besides this, the usability test also run to test the user friendly level of the system.

Last but not least, all the test result which showed positive result also included in this chapter.

Next chapter, the last chapter of this report will be discussing on the value of the system. The limitation and strength of the system as well as the contribution and the market price of the product to the society.

# CHAPTER VII

# CONCLUSION

## 7.1 Introduction

In this chapter, the final summary of this whole system will be concluded starting with elaborating the weakness and strength of the system. This can help to improve the user friendly to everyone regardless they are from technical background. We want to make sure the user interface can be easily performed by any type of user.



The feedbacks from the public are collected to improve the system. Some small and important improvement can be done in the short duration this semester, however, as for big issue and update will be bring forward for future improvement. The project is believed to bring advantage to the university and also to the public. The session also will be covered in this chapter. To ensure the advantage can be brought to the user, all the objectives stated in the chapter 1 must be achievable.

## 7.2 Observation on Weakness and Strength

There are few weakness found  in the system. Although the limitations found are not the serious problems, they do bring some inconvenience feeling to the user.

1. Total comments retrieved is only 50% number of what it got

   There is the only limitation of YouTube API. Due to some reasons, YouTube limited the total volume of video's comments up to only 50% of what it originally have. For example, Marques Brownlee latest video with titled "Samsung Galaxy Note 7 Impressions!" showing that there are 6,175 comments on the YouTube websites. However, the total comments that allowed being collected are approximately up to 3000 comments. There is the rule implied by YouTube, and it doesn't bring a huge impact when come to analytics and predictions part.

2. Long durations of data collection.

   The process of data collection can be taken up to 20 minutes to retrieve all search with containing 50 videos. Each video comes with its own description and information that highly needed to be part of the prediction process. The process turns to be so heavy when the average of 4000 comments for each video are also collected at the same time. Page loading time is obviously an important part of any website's user experience. The long durations of data collecting process will turn up killing the users. The user might quit the system anytime when losing their patience.



3. Lagging feel when data are retrieving

   As mentioned before, there will be a heavy load when data are retrieving to the system's database. A Large volume of data flowing through the narrow path will cause the traffic congestion and bad traffic will slow down the system. The process also included the sentiment analysis on those comments collected by analysis the polarity of the comments. This even increases the workload of processing. It is suggested to not using the system while data are retrieving. Go for bath or get yourself some foods and leave the system for 30 minutes.

4. Accuracy performance

   In term of accuracy, there is no way to know whether a prediction is correct. The system is running the predictions on Apple sale of 2016. As Apple haven't released their fourth quarter of financial report, the accuracy of the predictions will remain unknown. However, the predictions can always compare to the analysis result that produced by the bankers of wall street.

This project showing its strength as a powerful analytics tool. As a analytics tool, the data collected are most important component of the whole system.

1. The data are always up-to-date.

   As this is a real-time system, the data collected from YouTube API are always the latest data generated by YouTube that can also be viewed on YouTube website. YouTube has 17,000 channels with more than 100,000 subscribers which are significantly bringing impact on users. Every day, there are more than 300k of videos being uploaded to YouTube. As for this, obtaining the data that always up-to-date is an essential key for predictions.



2. Real world data

   The system consists of 2 kind of data, which are social media data and Apple iPhone historical sale data. Both data sharing the similarity where they are both real data. The key features of using real world data is they are all unpredictable. A lots of uncertainty can be found in the data where bring more interesting part to the system by discovering the secret hide inside the data.

3. Different perspectives of sale predictions.

   Research showed that many giant tech company carry out their own sale predictive algorithm based on historical sale data, company financial data and so on. By developing this system, might help in term of raising

consciousness of social media data. Using social media data as a source data for the predictive model, might be so accurate and precise. However, it can be always act as additional assistance tool to the current model to increase the accuracy from the perspectives of users.

Most of the feedback from others showing positive response, Enlarging the power of social network, and leveraging with the sale of the company is the key of success. Excellence observation are made by finding the relationships between social media and company sales. However, there are some negative feedback from the public stating that there are a lots of improvement need to made on the system. This including the accuracy of the result. Some also reported that retrieving social media data from single source is not convincing enough, it will be good to collect the data from 4 major social network, which including Twitter, Facebook, Instagram and YouTube.



**7.3 Propositions for improvements**

The improvements suggested to made to solve all the limitations and weakness mentioned in the part 1 of this chapter, except the for the constraint of YouTube API. Besides this, there are some enhancements that can be applied on data collection. More data to discover to increase the effectiveness of the system. In order to commercialize the product to the marker, system diverse for many type of mobile phone manufacturer and the system need to be fully automated to make this happen. Accommodating with new service which allowed user to get the company's historical data and reinforce this technique to the social media data. Last but not least, new algorithms for predicting need to be proposed to have better repercussion.

## 7.5 Project Contributions

Undoubtedly, this system will bring benefits to company when their products are being analyzed. However, some tuning also can be made to create new tool for university by adopting the same predictive model. The system can be developed to predict the popularity of UTeM by using all the social media data and calculate the number of student intake in the future.

## 7.5 Conclusion

As a conclusion, the system is not a best tool for analyze the social data as well as sale prediction. It does contains a lots of downside and limitations that need to be solved and improved. However, this project still considered as completed by achieving the goals and objectives that set in the proposal. The system help to bring inspirations to discover the impotency of social media data that cannot be neglected nowadays. By visualizing the data collected on the dashboard as well as the running a simple predictive algorithm to predict the sale. As the result of the prediction, Apple will failed to continue their greatest success in iPhone sales. In fact, this year around, Apple will have slightly lesser sales on iPhone 6s and iPhone SE compared to the sale of iPhone 6 in 2015.

# REFERENCES


Ioanăs, E. and Stoica, I., 2014. Social media and its impact on consumers behavior. International Journal of Economic Practices and Theories, 4(2), pp.295-303.

Vinerean, S., Cetina, I., Dumitrescu, L. and Tichindelean, M., 2013. The effects of social media marketing on online consumer behavior. International Journal of Business and Management, 8(14), p.66.

Gaikar, D . and Marakarkandy, B ., 2015. Product Sales Prediction Based on Sentiment Analysis Using Twitter Data. International Journal of Computer Science and Information Technologies, Vol. 6 (3)

Krishna, A., Zambreno, J. and Krishnan, S., 2013, December. Polarity trend analysis of public sentiment on YouTube. In Proceedings of the 19th International Conference on Management of Data (pp. 125-128). Computer Society of India.

Asur, S. and Huberman, B.A., 2010, August. Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ ACM International Conference on (Vol. 1, pp. 492-499). IEEE.

Cheng, X., Mehrdad, F., Ma, X., Zhang, C. and Liu, J., 2014. Understanding the YouTube partners and their data: Measurement and analysis. China Communications, 11(12), pp.26-34.

Lassen, N.B., Madsen, R. and Vatrapu, R., 2014, September. Predicting iphone sales from iphone tweets. In Enterprise Distributed Object Computing Conference (EDOC), 2014 IEEE 18th International (pp. 81-90). IEEE.

Chu, H., Tian, Y. and Yuan, H., Using Tweets for single stock price prediction.

Lin, S., Kong, X. and Yu, P.S., 2013, October. Predicting trends in social networks via dynamic activeness model. In Proceedings of the 22nd ACM international conference on Conference on information & knowledge management (pp. 1661-1666). ACM.

Johnson, P. (2012). MIT algorithm can predict which Twitter topics will trend. [online] ITworld. Available at: http://www.itworld.com/article/2717834/big-data/mit-algorithm-can-predict-which-twitter-topics-will-trend.html [Accessed 6 Jun. 2016].

Alba, D. and Greenberg, A. (2016). Foursquare Says Foot Traffic Can Predict iPhone Sales. [online] WIRED. Available at: http://www.wired.com/2015/09/foursquare-claims-foot-traffic-can-predict-weekend-iphone-sales/ [Accessed 8 Aug. 2016].

Edwards, J. (2016). Every major Apple analyst is now predicting a decline in iPhone sales - Business Insider. [online] Business Insider. Available at: http://www.businessinsider.my/apple-analyst-predicting-a-decline-in-iphone-sales-2016-1/?r=UK&IR=T#HQWuiRHgHEPltXKK.97 [Accessed 9 Aug. 2016].

Gayomali, C. (2014). One Way Google Could Use Skybox? To Spy On Apple's iPhone Supply Chain. [online] Fast Company. Available at: http://www.fastcompany.com/3032082/most-innovative-companies/one-way-google-could-use-skybox-to-spy-on-apples-iphone-supply-cha [Accessed 9 Aug. 2016].

Glueck, J. (2015). Right on Target: Foursquare Accurately Predicts iPhone Sales — Foursquare Direct. [online] Medium. Available at: https://medium.com/foursquare-direct/right-on-target-foursquare-accurately-predicts-iphone-sales-4140c450a396#.c1xq9jkp2 [Accessed 9 Jul. 2016].

Henry, V. (2011). Americans replace their cell phones every 2 years, Finns – every six, a study claims. [online] Phone Arena. Available at: http://www.phonearena.com/news/Americans-replace-their-cell-phones-every-2-years-Finns--every-six-a-study-claims_id20255 [Accessed 16 Jul. 2016].



Johnson, P. (2012). MIT algorithm can predict which Twitter topics will trend. [online] ITworld. Available at: http://www.itworld.com/article/2717834/big-data/mit-algorithm-can-predict-which-twitter-topics-will-trend.html [Accessed 6 Jun. 2016].

MIT News. (2012). Predicting what topics will trend on Twitter. [online] Available at: http://news.mit.edu/2012/predicting-twitter-trending-topics-1101 [Accessed 12 Jul. 2016].

Yu, S. and Kak, S., 2012. A survey of prediction using social media. arXiv preprint arXiv:1203.1647.

Liu, B., 2012. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), pp.1-167.

Mestyán, M., Yasseri, T. and Kertész, J., 2013. Early prediction of movie box office success based on Wikipedia activity big data. PloS one, 8(8), p.e71226.