

# Combinational Optimization of Camping Scheduling

Songsong Wang  
Shandong University of  
Science and Technology,  
Qingdao 266590, China

Xiaowen Ji  
Shandong University of  
Science and Technology,  
Qingdao 266590, China

Cong Liu\*  
Shandong University of  
Science and Technology,  
Qingdao 266590, China

---

**Abstract:** With the rise in popularity of river rafting, the park managers are faced with the problem of allowing more trips to travel down the river and providing wilderness experience for the visitors, with minimal contact with other groups of boats on the river as well. In this paper, the schedule of trips on the Big Long River (225 miles) is studied. The drifting vessels that can be chosen are oar-powered rubber rafts and motorized boats. By using the queuing theory, we propose several schedules, including the departure time, arriving time, duration (measured in nights on the river) and propulsion (motor or oar), based on some basic and practical assumptions. All schedules proposed here can provide visitors with a wilderness experience, with no two sets of campers occupying the same site at the same time. Based on the number of campsites, the transportation tool, the days of trip and some reasonable assumption, the paper model the above optimal schedule problem as a combinational optimization problem. We use heuristic algorithm to solve the above combinational optimization problem. First of all, we build the matrix which representing arrangement of a boat in 180 days. Secondly, we build the matrix which means the arrangement of campsites in one day. Third, we choose matrix which meet the constraint condition and have the maximized cardinal of line. The number of maximized cardinal of line is the optimal solution of the original combinational optimization problem. So we can see that the result is enormous, According to this model, we can get two-dimensional matrix, it will Simplify the results. The maximum of quantity of matrix which satisfy the conditions is optimal solution.

**Keywords:** matrix permutation and combination; Capacity of the river

---

## 1. DESCRIPTION

In this subject, the model is based on how to use optimized model to solve the boats of dock problem. We use the combinatorial optimization method [1], [2]. In this method, we suppose matrix to imitate the boats of dispatch. According to matrix addition to imitate the situation of camps. The daily situation of boats starting off can make up a two-dimensional matrix. Through a series of constraint conditions, we can optimize the result step by step.

There are  $Y$  camp sites on the Big Long River, distributed fairly uniformly throughout the river corridor. Given the rise in popularity of river rafting, the park managers have been asked to allow more trips to travel down the river. They want to determine how they might schedule an optimal mix of trips, of varying duration (measured in nights on the river) and propulsion (motor or oar) that will utilize the campsites in the best way possible. We also take many practical concerns into account. We discuss the influence of uncertain things on the trip that our algorithm is most useful in reduce the errors.

## 2. THE PROBLEM ANALYSIS

Our task is to design a model to calculate the number of the boats. Train of the thought: The opening days is 180 days. One boat need 6-18days, which is continuous, to pass through the river. So we should remove 6-18days continuous from the 180 days. And we can use a matrix  $M$  to express this situation. That is forming a matrix which includes 180 rows and  $Y$  columns. In every row only one element is 1 representing the campsite, with others is 0 representing that there is no campsite.

The biggest headaches are

- (1) How to arrange boats to make the  $X$  is maximal.
- (2) How to calculate the number of boats everyday.
- (3) How to choose boats and how many boats should be sent in one day,
- (4) How to select the campsites and satisfied the constraint condition.

First of all, we should build a matrix to show arrangement of a boat in 180 days. Secondly, we should build a matrix to show arrangement of campsites in one day. Third, we should

choose matrix which meet the constraint condition and to do permutations and combinations. Also, we must be clear that we have constraints, which include some factors, such as time, the distance between two campsites adjacent, the speed of the boat. So we should exclude some unsuited plans.

### 3. ASSUMPTIONS

Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

(1) A boat can stay at a campsite but once a day. Because the government agency responsible for managing this river wants every trip to enjoy a wilderness experience, with minimal contact with other groups of boats on the river. The government agency responsible for managing this river wants every trip to enjoy a wilderness experience, with minimal contact with other groups of boats on the river.

(2) All the boats have the same departure time in the morning, doing this we can make sure that when the later boat arrive at the campsite the campsite is empty.

(3) We should suppose the weather of everyday is good, so each boat can travel at its appointed speed.

(4) The number of starting boats (or landing boats) is affected by one factor, which is the number of campsites.

(5) The time of starting boat and stopping time is ignored, only in this way, can we make the model more simple.

(6) Every campsite can hold only one boat at the same time, considering that the campsites is not big enough to hold more than one boat.

(7) The boats must be at the campsites before 6:00 pm. Because it is safer for visitors to stay at campsites at night. And visitors can have a rest and prepare equipment for tomorrow. What's more, the river is nature and valuable, we should not use it frequently.

(8) Every boats have their own independence at any time. so visitors can enjoy themselves.

(9) The days every boats stay at one campsite can not exceed one.

(10) Every trip can not change its transportation during the journey. Because managers have make the schedule, so visitors need to make an appointment on the phone.

### 4. MODEL

Suppose the number of campsites is  $Y$ , and number the campsites as  $1, 2, \dots, Y$  along the river from upstream to downstream. We can use a  $Y$ - dimension vector to denote the camping arrangement of a group at some night on the travel, the element of vector is 0 or 1, for example  $(1,0,0,\dots,0)$  represent this group will camp at campsite 1. If the element of the vector is 0, this means that the group does not start or have finished his travel. So we can use a matrix  $M_i$  represent a travel group  $i$  travel arrangement during a six month period, which is a  $180 \times |Y|$  matrix, the row vector is the camp arrangement in some night. Suppose the set of feasible travel arrangement is  $L=\{1,2,\dots,n\}$ . Because two sets of campers can occupy the same site at the same time. This means that we should find a subset of  $L$  such that two sets of campers can occupy the same site at the same time. On the other hand, the river managers want to utilize the campsites in the best way possible. This means the cardinal number of the subset must be maximized. So we can model the river manager's problem as an optimization problem as follows:

$$\begin{aligned} & \max |U| \\ & s.t. \\ & \begin{cases} U = \{i_1, \dots, i_m\} \in P(L), m = 1, \dots, n \\ (M_{i_1}, \dots, M_{i_m}) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \leq \begin{pmatrix} (1, \dots, 1) \\ \vdots \\ (1, \dots, 1) \end{pmatrix} \end{cases} \end{aligned}$$

In the ideal case, all the boats fully comply with the arrangement of makers. Campsite can be built everywhere along the river which is 225miles long. So the campsites can be distributed fairly uniformly throughout the river corridor.

Suppose the number of campsites is  $Y$ ,  $Z_i$  represent the  $i$  campsites.

$$z_i = \begin{cases} 1 \\ 0 \end{cases} \quad (1 \leq i \leq Y, \quad i \in N) \quad (1)$$

Here 1 represents that the boats will call in at  $i$ , and 0 represents that the boat does not stop at the place. Then this case can be expressed by a one-dimensional vector  $T$ .

$$T = \{z_1, z_2, \dots, z_Y\}$$

The daily situation of boats starting off can make up a two-

dimensional matrix  $A = (a_{ij})_{m \times n}$  ( $a_{ij} = 0$  or  $1$ )

$a_{ij}$  represents the situation of the  $i$  boat stopping at the  $j$  campsite.

Suppose  $p$  represents a one-dimensional vector with all of its element is 1. And the arrangement of the boats must under this constraints

$$\begin{aligned} \sum_{i=1}^m p a_{i1} &\leq 1 \\ \sum_{i=1}^m p a_{i2} &\leq 1 \\ \sum_{i=1}^m p a_{im} &\leq 1 \end{aligned} \quad (2)$$

Meanwhile, the daily number of boats starting off should be record.

The arrangement of boats ought to make the utilization ratio of campsites as high as possible and reach the highest number of boats starting off in this 180 days.

The model should be built when the target of system is the best.

$$\max X = \sum_{m=1}^{180} x_m \quad (3)$$

$x_m$  represents the daily number of boats setting off the original.

Overall optimization;

Train of the thought: The opening days is 180 days. One boat need 6-18 days, which is continuous, to pass through the river. So we should remove 6-18 days continuous from the 180 days. And we can use a matrix  $M$  to express this

## 5. SOLUTION AND RESULT

(1) The first step: Supposing the value of  $Y$

If we don't assign a value to  $Y$  and we don't optimize our choosing, we will have a great many of results.

situation. That is forming a matrix which includes 180 rows and  $Y$  columns. In every row only one element is 1 representing the campsite, with others is 0 representing that there is no campsite. Then using matrix multiplication or matrix logic exclusive OR (XOR) to express the constraint situation. Everyday the boats can not stay at the same campsite. First, we should list all the matrix ( $m$ ) satisfying the constraints. Then we combine this matrix ( $m$ ) satisfying the constraints. The number of  $m$  in the combination which has the most number of matrix is  $x$  which is what we need.

Such as:

$$\begin{aligned} (M_{ij})_{180 \times Y} &= \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ \sum_{j=1}^Y M_{ij} &\leq 1 \quad (1 \leq i \leq 180) \end{aligned} \quad (4)$$

that is every row has only 1 or no 1.

The 6-8 days continuous which are selected from the 180 days make up the combinations. And the selection of the first campsite is in a certain range.

Supposing that the first campsite of first day the boat stop at is the  $j$  campsite and the boats move ahead as fast as possible

$$1 \leq j < (8 \times 12) \div (225 \div Y) \quad (5)$$

And the next situation of 1 should be bigger than the value of  $j$ .  $M_i$  is the situation the  $j$  boat select.

$$\text{Provided } N_{1t} = \begin{Bmatrix} M_1 \\ M_2 \\ M_3 \\ \dots \\ M_t \end{Bmatrix}$$

Per element of the one-dimensional vector represent the situation of arrangement. Supposing that there  $L$  answers which may solve the problem.

The best answer  $N \in L$ .

From the result of software, when  $Y=18$  we can get different plans of stopping the boats just as the days on the trip is 6. Here we assume all the boats are fast boats.

**Table 2: the whole boats' camp arrange**

Groups days	1	2	3	4	5	6	7	.....	180
1	1	1	1	1	1	1	0	00000	0
2	0	1	1	1	1	1	1	00000	0
3	0	0	1	1	1	1	1	10000	0
4	0	0	0	1	1	1	1	11000	0
.....	...	..	..	..	..	..	..	..	..

The situation of the boats setting off per day according to one of the groups is as follows:

The average distance the boats travel is 37.5 miles per day, the average time the boats travel is 4.6875 hours. It will take 1.5625 hours to run across the two campsites which are adjacent. Everyday one boat can travel as many as 12 hours ,that is 96 miles. Such as: When a boat pass across 7 campsites a day, we have to select 5 campsites from the last 11 campsites.

For example :1-7,2-8,3-9,4-10,5-11,6-12,7-13, 8-14,9-15,10-16,11-17,12-18

$$M_1 = \{111110000001000000\}$$

$$M_2 = \{01010101011001000000\}$$

$$M_3 = \{00100101011011000000\}$$

$$M_3 = \{01010101010011000000\}$$

(2) Then get the situations when a boat pass across 6, 5,4,3campsites.

From the information we can know the running conditions of the Big Long River;

**Table 3: different date of percentage**

Trip choosing	Time choosing	percentage
Short motor trips	6–8 days	22
Long motor trips	9 or more days	32
Short oar trips	14 or fewer days	26
Long oar trips	15 or more days	29

In order to directly received through the senses, we draw the pie chart as follows.

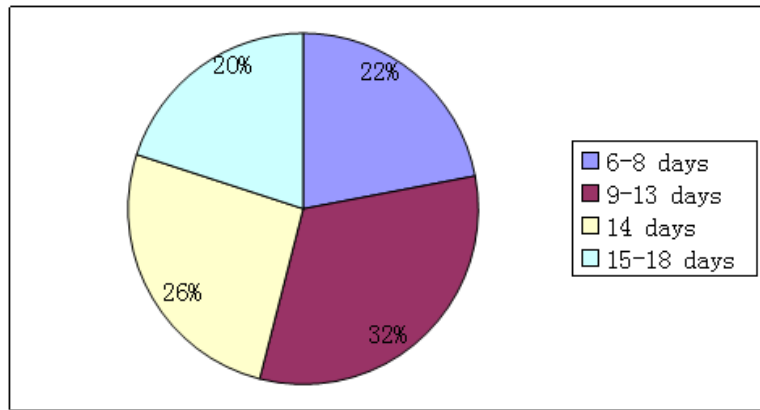


Figure 3: different date of percentage (pie chart)

We can get a method which can distribute the time on average. In order to get full use of the camps, we can fix short-time with long-time.

Table 4: average distribution of boats in every day

	the distribution of boats in 13 days												
one	6	7	8	9	10	11	12	13	14	15	16	17	18
two	18	17	16	15	14	13	12	11	10	9	8	7	6
three	6	7	8	9	10	11	12	13	14	15	16	17	18
four	15	14	13	12	11	10	9	8	7	6			
sum	45	45	45	45	45	45	45	45	45	45	40	41	42

## 6. ANALYSIS OF RESULT

Y=45, there are 13 different plans for boats

If we don't think about the constraint condition, we will have many results. For example:

```

00000001000000010000000100000001000000010000
00000010000001000000100000010000001000000100
00000100000100000100000100000100000100000100
00000100000100000100000100000100000100000100
00001000010000100001000010000100001000010000
00001000010000100001000010000100001000010000
00010001000100010001000100010001000100010001
00010001000100010001000100010001000100010001
00010001000100010001000100010001000100010001
00010001000100010001000100010001000100010001
00100100100100100100100100100100100100100100
00100100100100100100100100100100100100100100
    
```

00100100100100100100100100100100100100100100100100100

So we can see that the result is enormous.

In local optimum, we use a two dimension matrix to express the situation of boats every day. So we can get the max number of combination as the max quantity of boats. And by this analogy, we can calculate the sum of the boats in 180 days. In the global optimization, we use matrix M to signify driving conditions in 180 days. The element of the matrix

### 7. BACKGROUND

In nowadays more people like to travel and meantime the river trips is very popular in recent years. Because river trips is an adventure sports. Visitors to the Big Long River (225 miles) can enjoy scenic views and exciting white water rapids. The river is inaccessible to hikers, so the only way to enjoy it is to take a river trip that requires several

show the situation of boats. The maximum of quantity of matrix which satisfy the conditions is optimal solution. Because of the time ,in the solving of the model ,we only deal with one situation ,and other situations are similar with the situation. In order to make the computing more simple, we set Y as 18 .Meanwhile ,all the boats are the fast boats and all the boats need 6 days to pass across the river.

days of camping. In order to make the journey more safe we have to build some campsites along the river.

We know that the trip involves so many factors that we can not calculate the accurate the sum of quantity of boats in one year ,so we ought to simplify the travel. In order to make more boat trips be added to the Big Long River’s rafting season ,we design an optimized model to solve this problem.

**Table 1: Popular attraction sites in the river corridor during the typical flow regimes of 1998-2000(sheet)**

River mile	Site name	Rank	Frequency
136.2	Deer Creek	1	0.91
156.8	Havasu	2	0.88
32.8	Redwall	3	0.82
116.5	Elves Chasm	4	0.77
61.4	Little Colorado	5	0.66
87.8	Phantom	6	0.61
108.7	Shinumo	7	0.56
147.9	Matkatamiba	8	0.48
132.0	Stone CK	9	0.41
47.2	UPR Saddle	10	0.34
31.6	South Canyon	11	0.31
53.0	Nankoweap	12	0.3
84.0	Clear CK	13	0.28
34.8	Nautiloid	14	0.26
133.8	Tapeats	15	0.25

In order to directly run over the senses, we draw the histogram as follows.

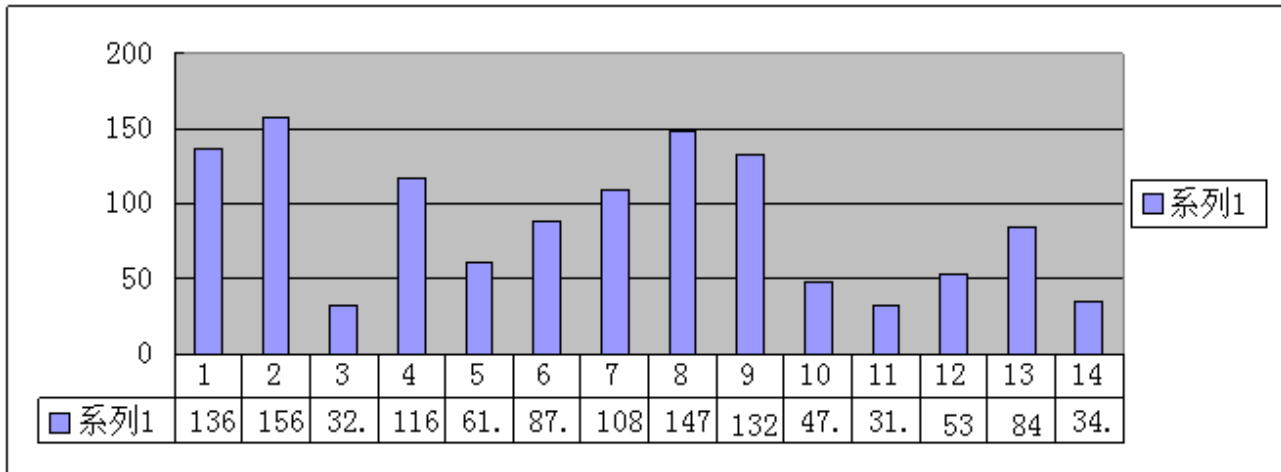


Figure 1: Popular attraction sites in the river corridor during the typical flow regimes of 1998-2000(histogram)

So we can know the generally range of Y from the chart above. When we do model analysis and mode solution, we will have a reasonable assumption. When we figure out one of the X, we will compare with the quantity of the camps above. If the result of assumption is far away from it. we will adopt another one.

The dates can regards as one part of one row of matrix  $M_{ij}$ . The general methods to solve the combinatorial optimization problems of 0 - 1 are method of exhaustion and implicit enumeration. The method of exhaustion need to test all the solutions ,provided the model has K variables , then we need to examine  $2^Y$  times, the calculated amount is very big. The implicit enumeration impliedly search all the solutions, and

this method can be divided into Backtracking Algorithm ,Branch and Bound method, and Sequential Combination Tree Algorithm, in this way ,we can reduce the calculated amount , and find the best answer. According to the feature of this model ,this article determine the search range by the methods of solution space decomposition and of solution space limit method, and seek the best answer by the method of Sequential Combination Tree Algorithm. From the picture ,we divide the solution space into several subspaces according to certain rules. And confine the subspaces which are infeasible and which are feasible but are not the best subspaces, so as to narrow down the search range .In the new search range ,we can calculate the best answer exactly.

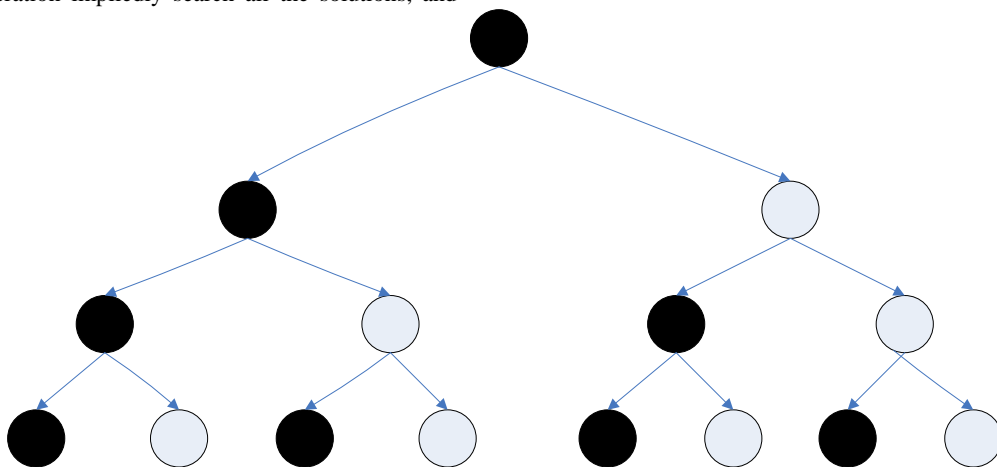


Figure 4:simulation of choosing in every day(the black round show the camp is full, the white round show the camp is empty)

The black dots represent 1 ,the white dots represent 0,and the layer where the dots locate represents the situation of the campsite .There are Y layers. Supposing the days the sailing

$$\sum_{i=1}^Y z_i = d$$

time of the boats is d days.  $6 \leq d \leq 18$  then

## 8. ADVANTAGE AND DISADVANTAGE

The structure of Heuristic algorithm is based on intuitive judgment or experience. We can get a feasible solution of the problem in the condition of acceptable cost (time cost, space cost etc.)The rate of deviation between the solution that we get and the optimum solution can not be predicted. We can get the best solution under the acceptable conditions. but we can

not guarantee the solution is the optimal solution and it might be infeasible. Even in most cases. we can not point out the approximate between the optimal solution and the solution we get.

Although we not always can get optimal solution ,but we can get the solution most close to the optimal solution.

Heuristic supply low bifurcate rate for each to solve specific problems of search trees each node, so they have better efficiency of computing .It is easier to get reasonable solution. Comparing these solution to choose the solution as optimum solutions.

Our model effectively achieved all of the goals we set initially. It was fast and could handle large quantities of data, but also had the flexibility we desired. Though we did not test all possibilities, we showed that our model optimizes state districts for any of a number of variables. If we had chosen to input more income, we could have produced high-quality results with virtually no added difficulty. As well, our method was robust.

## 9. CONCLUSIONS

During the process we realized that teamwork is very important in our daily life ,especially when the mission can not be handled. by only one person. Before action we should divide the job into different parts .Also we need to make sure that all of the members of the team can have his own duty .Only every one in the team make his great effort ,can the team achieve great grades. We should consider every problem from all the aspects .In the process we should collect a mass of material ,so effective information search is of very important. As to the materials collected ,it is vital to refine the information ,which is useful for us. In the future, we prefer to integrate the optimization with some traditional data mining approaches [3], [4], [5] to find more effective ways for optimization.

## 10. REFERENCES

- [1] Sherali H D, Desai J, Rakha H. A discrete optimization approach for locating automatic vehicle identification readers for the provision of roadway travel times[J]. Transportation Research Part B: Methodological, 2006, 40(10): 857-871.
- [2] Lin X, Shroff N B. Utility maximization for communication networks with multipath routing[J]. IEEE Transactions on Automatic Control, 2006, 51(5): 766-781.
- [3] Ting Lu, Qi Gao, Xudong Wang, Cong Liu, “Modified Page Rank Model in Investigation of Criminal Gang”, International Journal of Science and Engineering Applications, 4(3): 100-104, 2015.
- [4] Baohua Liu, Xudong Wang, Qi Gao, Cong Liu, “A K-Means based Model towards Ebola Virus Prorogation Prediction”, International Journal of Science and Engineering Applications, 4(3): 95-99, 2015.
- [5] Libao Zhang, Faming Lu, An Liu, Pingping Guo and Cong Liu\*, “Application of K-Means Clustering Algorithm for Classification of NBA Guards”, International Journal of Science and Engineering Applications, 5(1), pp. 1-6, 2016.



# Evolution and Influence Measurement Association Information Network

Wenchao Li  
Shandong University of  
Science and Technology,  
Qingdao 266590, China

Jijuan Jiang  
Shandong University of  
Science and Technology,  
Qingdao 266590, China

Cong Liu\*  
Shandong University of  
Science and Technology,  
Qingdao 266590, China

---

**Abstract:** This paper mainly using SIR model, based on least squares polynomial fitting GM (1,1) model, Information flow reasonable quantitative analysis and evaluation of the impact of information and the public interest and public opinion to make a qualitative analysis. First, the study of journalism and communication problems model. Using SIR model divided population into three categories: Communicators, unknown person and rem. Then contra the news spread found the SIR model. Through qualitative analysis get rationality of SIR model. Then get specific data unknown persons in accordance with the SIR model was built using the least squares polynomial fitting based on those data, the unknown polynomial fitting. If the data is in line with the model it can be considered as news data, otherwise, the data is not news. Besides, to validate the reliability of model and to forecast the trend of the information flows today. Getting a group of reliable news media data through the local statistical offices personnel is consistent with the model. Since there is no question as to the data to obtain information dissemination five times, so this article by setting the initial value of SIR model, Since there is no question as to get information dissemination five periods of data, so this article by setting the initial value of SIR model to obtain information dissemination five periods of data, and then use the GM (1,1) model predict unknown persons data for each day, obtained dissemination of information specific data of each day. Whereby the trend today of unknown people the number of information dissemination, through testing the GM (1,1) model, it showed that the use of GM (1,1) model to predict today's information communication situation is entirely possible.

**Keywords:** SIR model; least squares polynomial fitting; GM (1,1) model

---

## 1. DESCRIPTION

Information is spread quickly in today's tech-connected communications network; sometimes it is due to the inherent value of the information itself and other times it is due to the information finding, its way to influential or central network nodes that accelerate its spread through social media. While content has varied -- in the 1800s news was more about local events (e.g., weddings, storms, deaths) rather than viral videos of cats or social lives of entertainers -- the prevailing premise is that this cultural characteristic to share information (both serious and trivial) has always been there. However, the flow of information has never been as easy or wide-ranging as it is today, allowing news of various levels of importance to spread quickly across the globe in our tech connected world. By taking a historical perspective of flow of information relative to inherent value of information, the Institute of Communication Media (ICM) [1], [2] seeks to understand the evolution of the methodology, purpose, and functionality of society's networks. Specifically, your team as part of ICM's Information Analytics Division has been assigned to analyze the relationship between speed/flow of information vs inherent value of information based on consideration of 5 periods: in the 1870s, when newspapers were delivered by trains and stories were passed by telegraph; in the 1920s, when radios became a more common household item; in the 1970s, when televisions were in most homes; in the 1990s, when households began connecting to the early internet; in the 2010s, when we can carry a connection to the world on our phones. Your supervisor reminds you to be sure to report the assumptions you make and the data you use to build your models.

The problems to be solved are:

- 1) Develop one or more model(s) that allow(s) you to explore the flow of information and filter or find what qualifies as news.
- 2) Validate your model's reliability by using data from the past and the prediction capability of your model to predict the information communication situation for today and compare that with today's reality.

## 2. SYMBOL CONVENTIONS AND ASSUMPTIONS

### 2.1 Symbol Conventions

$s(t)$ : This news did not know who (unknown persons) ratio in the population;

$i(t)$ : People know this and to spread the news (communicators) the proportion in the population;

$r(t)$ : People do not know but this news spread (out of those) the proportion in the population;

$\lambda$ : Day transmission rate communicators;

$\mu$ : Spread spread their number;

$\sigma$ : Day stop transmission rates.

### 2.2 Conditions Model assumptions

Due to constraints time and our limited ability level, in order to ensure the reasonableness of rationality of the model and the model of the flow of information, the following assumptions are made.

1) In order to simplify the problem, assume that people were divided into three categories: unknown, disseminators and were removed.

2) Takes into account the short time the news spread, so the news media during the investigation of the total number of regions  $N$  unconsidered a birth or death, do not consider migration, and in days of measurement.

### 3. ANALYSIS OF THE PROBLEM

For question 1, develop one or more models, allowing you to explore the flow of information, filtering, or can find content that can be the news. Due to the spread of infectious diseases is similar to flow of information, so this paper intends to model the flow of information use the SIR Epidemic Model, by setting the parameters of the first test to obtain a model of the flow of information, through qualitative analysis, to test model is reliable. Then depending on the model obtain the data, this paper using the least squares polynomial fitting to fit the flow of information polynomial model. If the data obtained are consistent with the model, it can be considered the given data as news data, otherwise, the data is not news.

For question 2, using past data verify the reliability of the model; forecast the situation of today's information communication, verify the ability of prediction, for comparison with today's reality. This paper by looking for other data get reliable information, then obtained the resulting data about the time of trend graph, and compared with in the model image of Question 1, if reliable data is consistent with the model, can be obtained the model with a certain reliability. By looking for information to give the title to the information flow in the five periods, and this paper intends to use GM (1,1) model based on specific data of five periods' information flows, obtained communication skills today, and compare with today's reality.

## 4. MODELS

### 4.1 Establishment of the News' Model

The problem to be solved is: Develop one or more models allowing you to explore the flow of information, filtering, or can find content that can be the news. Due to the spread of infectious diseases is similar to flow of information, so this paper intends to model the flow of information use the SIR Epidemic Model, by setting the parameters of the first test to obtain a model of the flow of information, through qualitative analysis, to test model is reliable. Then depending on the model obtain the data, this paper using the least squares polynomial fitting to fit the flow of information polynomial model. If the data obtained are consistent with the model, it can be considered the given data as news data, otherwise, the data is not news.

#### 4.1.1 SIR model on the news

Summary of population represented by  $N$ , Model assumptions are as follows:

1) This article will be divided into three groups,

a. Person does not know this news (unknown persons),  $s(t)$  represents an unknown proportion in the population;

b. People who know and spread the news (communicators),  $i(t)$  indicates the proportion of communicators in the crowd;

c. People who know the news, but does not spread (out of person),  $r(t)$  shows proportion of out of person in the population.

2) Day spread rate of communicators is  $\lambda$ , day stop transmission rate is  $\mu$ . Their dissemination number of Disseminate is  $\sigma = \lambda / \mu$ .

3) Taking into account that the news spread in a short time, therefore during area the total number of news media in the study is the constant  $N$ , that is not considered either birth or death, nor migrating, and measured in days.

Under the above scenario, crowd flow chart is as follows:



Figure. 1 Crowd flow chart

The assumption is clearly:

$$\begin{aligned} s(t) + i(t) + r(t) &= 1 \\ N \frac{ds}{dt} &= -\lambda N s i \\ N \frac{di}{dt} &= \lambda N s i - \mu N i \\ N \frac{dr}{dt} &= \mu N i \end{aligned}$$

The proportion of unknown person and communicators in the initial time are  $s_0(s_0 > 0)$  and  $i_0(i_0 > 0)$  (here assume the initial value of the removed persons is  $r_0 = 0$ ), So get SIR model initial value problem are as follows

$$\begin{cases} \frac{ds}{dt} = -\lambda s i, & s(0) = s_0 \\ \frac{di}{dt} = \lambda s i - \mu i, & i(0) = i_0 \\ \frac{dr}{dt} = \mu i, & r(0) = 0 \end{cases}$$

And then have  $s + i + r = 1$  get  $dr / dt = -di / dt - ds / dt$ , thus the third formula equation becomes identity, thereby simplifying the model as follows

$$\begin{cases} \frac{ds}{dt} = -\lambda s i, & s(0) = s_0 \\ \frac{di}{dt} = \lambda s i - \mu i, & i(0) = i_0 \end{cases}$$

Above initial value problem can not get analytical solution, only be obtained by numerical solution to get numerical solution.

For example, take  $\lambda = 1, \mu = 0.3, i(0) = 0.02, s(0) = 0.98$ , then get numerical solution is obtained in the following table,  $i(t), s(t)$  and  $i$ - $s$  curve as shown below correspondingly.

**Table 1.  $i(t)$  and  $s(t)$  numerical solution of SIR model**

t	0	1	2	3	4	5	6	7	8
i(t)	0.02	0.039	0.0732	0.1285	0.2033	0.2795	0.3312	0.3444	0.3247
s(t)	0.98	0.9525	0.9019	0.8169	0.6927	0.5438	0.3995	0.2839	0.2027
t	9	10	15	20	25	30	35	40	45
i(t)	0.2863	0.2418	0.0787	0.0223	0.0061	0.0017	0.0005	0.0001	0
s(t)	0.1493	0.1145	0.0543	0.0434	0.0408	0.0401	0.0399	0.0399	0.0398

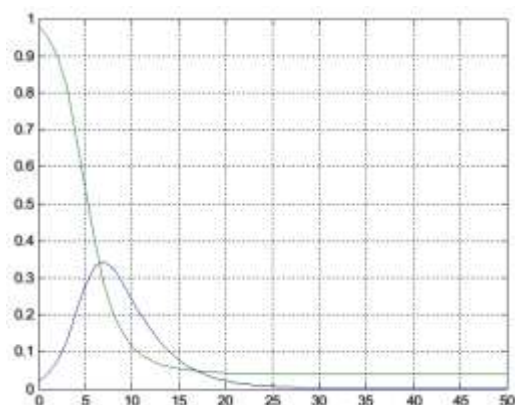


Figure. 2  $i(t)$  and  $s(t)$  curve of SIR model

From the perspective of qualitative analysis, the results of this model has been broadly in line with the trend of the time in real life people of a news attention. In real life, when a news just happened, people's attention will increase sharply this news, then quickly reached a peak, and then the news will concern a sharp decline, then forgot this news slowly. The unknown number of this news will be a sharp reduction in a very short period of time, but there is always a small group of people learn the news because of this reason or another. From the perspective of qualitative analysis point of view, SIR model established in this paper according to the propagation of news.

#### 4.1.2 Polynomial fitting based on least squares

As mentioned above, from the perspective of qualitative analysis, it can be obtained: the dissemination of information relating to the establishment of the above-SIR propagation model in line with real-life news. Therefore, this article will based on the data obtained by above example, to be a polynomial fitting based on least squares information data for the news data.

##### 4.1.2.1 polynomial fitting model based on least squares

Curve fitting problem formulation is, given a set of (two-dimensional) data, namely  $n$  point on the plane  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ ,  $x_i$  different from each other, seek a function (curve)  $y = f(x)$ ,  $f(x)$  under certain guidelines so that all the data points closest to that is the best fit curve.

Linear least squares is the most common method to solve curve fitting, the basic idea is

$$f(x) = a_1 r_1(x) + a_2 r_2(x) + \dots + a_m r_m(x)$$

Where  $r_k(x)$  is a pre-selected set of linearly independent functions,  $a_k$  is undetermined coefficients ( $k = 1, 2, \dots, m, m < n$ ).

Guideline is make square of distance  $\delta_i$  that  $y_i, i = 1, 2, \dots, n$ , and  $f(x_i)$  and with a minimum, known as a A least squares criterion.

1) Determine the coefficient  $a_k$

$J(a_1, \dots, a_m) = \sum_{i=1}^n \delta_i^2 = \sum_{i=1}^n [f(x_i) - y_i]^2$  note for the sake of the  $a_1, \dots, a_m$  to a minimum. Simply use Extreme necessary conditions

$$\frac{\partial J}{\partial a_k} = 0 (k = 1, 2, \dots, m)$$

obtained the linear equations on  $a_1, \dots, a_m$

$$\sum_{i=1}^n r_j(x_i) \left[ \sum_{k=1}^m a_k r_k(x_i) - y_i \right] = 0, (j = 1, 2, \dots, m)$$

which is

$$\sum_{k=1}^m a_k \left[ \sum_{i=1}^n r_j(x_i) r_k(x_i) \right] = \sum_{i=1}^n r_j(x_i) y_i, (j = 1, 2, \dots, m)$$

$$R = \begin{bmatrix} r_1(x_1) & \dots & r_m(x_1) \\ \vdots & & \vdots \\ r_1(x_n) & \dots & r_m(x_n) \end{bmatrix}_{n \times m},$$

$$A = [a_1, \dots, a_m]^T, Y = [y_1, \dots, y_n]^T$$

From above, the equation can be expressed as  $R^T R A = R^T Y$ , When  $\{r_1(x), \dots, r_m(x)\}$  linearly independent,  $R$  full column rank,  $R^T R$  reversible. So

$$A = (R^T R)^{-1} R^T Y$$

2) selection of  $r_k(x)$  function

Faced with set of data  $(x_i, y_i), i = 1, 2, \dots, n$ , use linear least squares curve fitting, first and foremost, but also the crucial step is to appropriately select  $r_1(x) \dots r_m(x)$ . If through mechanism analysis, should be able to know what kind of function between  $y$  and  $x$ , it is easy to determine  $r_1(x) \dots r_m(x)$ . Unable to know the relationship between  $x$  and  $y$ . Typically can plot the data  $(x_i, y_i), i = 1, 2, \dots, n$ , visually determine what kind of curve fitting them should use. People often use a curve

(i) a straight line

$$y = a_1x + a_2$$

(ii) the polynomial

$$y = a_1x^m + \dots + a_mx + a_{m+1}$$

(generally,  $m=2,3$ , not too high)

(iii) hyperbola (only one)

$$y = \frac{a_1}{x} + a_2$$

(iv) an exponential curve

$$y = a_1e^{a_2x}$$

For the exponential curve, the need for variable substitution before fitting into the linear function  $a_1, a_2$ .

Given a set of data, what kind of curve fitting the best, possible on the basis of intuitive judgments to elect several curves, then compare, behold Article *J* of the smallest least squares curve index.

### 3) Least Squares

In unconstrained optimization problems, some important special cases, for example, the objective function by the square of the number of functions and composition. Such functions can generally be written as:

$$F(x) = \sum_{i=1}^m f_i^2(x), x \in R^n$$

Where  $x = (x_1, \dots, x_n)^T$  general assumptions  $m \geq n$ . We minimization problem of the kind of function:

$$\min F(x) = \sum_{i=1}^m f_i^2(x)$$

called least-squares optimization problem.

**4.1.2.2 The least squares polynomial fitting of News**  
 Based on data obtained using the Matlab program of above SIR model. The unknown proportion of the number of time-varying polynomial equation:

First Order:  $s(t) = -0.0188t + 0.6284$

Second Order:  $s(t) = 0.0013t^2 - 0.0716t + 0.8900$

Third Order:  $s(t) = -0.0001t^3 + 0.0052t^2 - 0.1355t + 1.0733$

Fourth Order:  $s(t) = 0.0000t^4 - 0.0002t^3 + 0.0080t^2 - 0.1593t + 1.1134$

As can be seen, in this fourth order polynomial coefficient of the highest degree term  $s(t)$  is close to zero, so in this pair  $s(t)$  only polynomial fitting to Fourth - order polynomial. And the

minimum square error form one to four order shown in Table 2.

**Table 2. least squares polynomial fitting difference of  $s(t)$**

Order Number	First Order	Second Order	Third Order	Fourth Order
The minimum square error	0.9956	0.3253	0.0811	0.0673

As can be seen from Table 2,  $s(t)$  least squares polynomial fitting difference from the Third Order Fourth Order are just 0.0138, and the fourth-order polynomial coefficients of the highest order term  $s(t)$  is close to zero. Therefore, in this select third-order polynomial fitting of  $s(t)$ , so the fitting function of  $s(t)$  is

$$s(t) = -0.0001t^3 + 0.0052t^2 - 0.1355t + 1.0733$$

The image of fitting function and the image of specific data are shown in Figure 3.

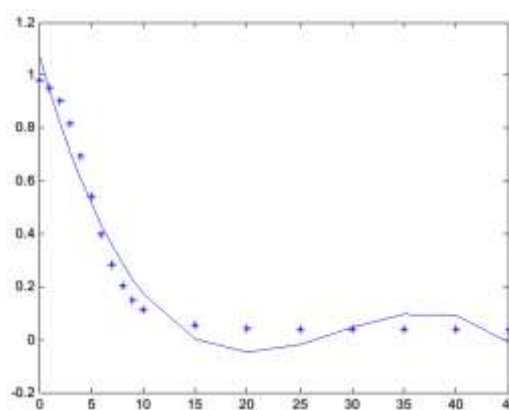


Figure. 3 The image of fitting function and the image of specific data

### 4.1.3 Data determined in the news

As mentioned above, from the perspective of qualitative analysis, it can be obtained, the dissemination of information relating to the establishment of the above-SIR propagation model in line with real-life news. And the data obtained in the SIR model of news spread performed function fitting. Therefore, the data was subjected to qualitative analysis( its scatter plot, unknown number with respect to time trends, etc.),and performed polynomial fitting to this set of data .If this set of data consistent with the established model can be considered that it is news, otherwise the message is not news.

## 4.2 Validation of model reliability and prediction information propagation today

### 4.2.1 Verify the reliability of the model

Linked to a relative working in local statistical offices by their families, Since people get a local news dissemination of data, and use this data to verify the reliability of the model herein, Although only we got a set of reliable data, data is little, but there are certain confirmatory. The proportion of the number of unknown data obtained as shown in Table 3.

**Table 3. a local news unknown proportion of data**

day	0	1	2	3	4	5	6	7	8	9
unknown people	0.98	0.95	0.88	0.76	0.58	0.39	0.25	0.15	0.10	0.07
day	10	11	12	13	14	15	16	17	18	19
unknown people	0.05	0.04	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01
day	20	21	22	23	24	25	26	27	28	29
unknown people	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
day	30	31	32	33	34	35	36	37	38	39
unknown people	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

A scatter plot local news by unknown proportion of the data shown in Figure 4.

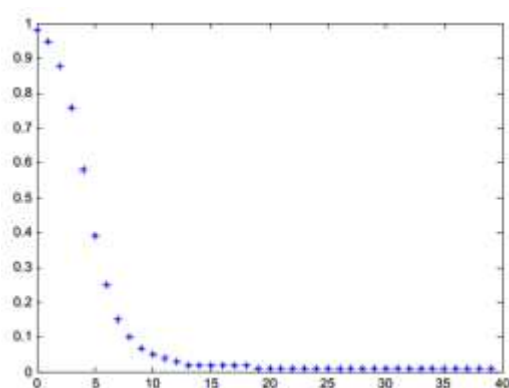


Figure. 4 Scatter plot in a local news unknown proportion of data

As can be seen from the figure, the number of unknown proportion of data that reliable data trends over time is almost the same with the image built SIR model above, to some extent, this article can explain the model has a certain reliability.

#### 4.2.2 Based on GM (1,1) model to predict the spread of today's information

##### 4.2.2.1 GM (1,1) basic ideas and principles

Gray theory think that every random amount is in a certain range, a certain period of time changes the amount of gray and gray process, incomplete of system information , uncertainty is the greyness. Do not look for their data processing and statistical probability distribution law, but processing the raw data for certain, It becomes a regular time-series data, found mathematical model based on this. Using GM (1,1) model based on the cumulative number of generated columns, model works as follows:

1) a cumulative for raw data

Let the original gray data  $x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)$ ,

denoted  $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$

Make a cumulative yield, get

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)), (k = 1, 2, \dots, n)$$

among them

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$$

Cumulative number of columns overcome volatility and randomness of raw data , change into strong regularity increasing the number of columns, be ready to build predictive models in the form of differential equations .

2) establish GM (1,1) model

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

Differential the equation is the gray forecast model GM (1,1),  $a, u$  which is a constant, it can be obtained by least squares fitting

$$\begin{pmatrix} a \\ u \end{pmatrix} = (B^T B)^{-1} B^T Y_n$$

among them

$$B = \begin{pmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{pmatrix}$$

$$Y_n = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))$$

Differential  $\frac{dx^{(1)}}{dt} + ax^{(1)} = u$  solutions (referred to as time response function) is

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a}$$

Formula prediction the formula is the number of columns, Because the formula is predicted value once accumulated generating sequence can be original series value determined by the following formula

$$\hat{x}^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$$

##### 4.2.2.2 Based applications GM (1,1) Model

As the group members did not find the specific data of five period about the situation of communicate news and information, therefore we decided to use the SIR model established above, by changing its initial value, to get information communication situation of five times specific data . Although this article does not use reliable data, but presents a solution to the problem. The initial value of the group on the basis of relevant experts on the use of the five periods as shown in Table 4

**Table 4. The initial value of the SIR model in Table 3 of the five periods**

period	1870s	1920s	1970s	1990s	2000s
$\lambda$	0.5	0.7	0.9	1.1	1.3
$\mu$	0.2	0.25	0.3	0.35	0.4

According to SIR model, using Matlab software program to obtain five periods of time the number of unknown persons  $s(t)$  and communicators  $i(t)$  with the proportion of time trends. as shown in figure 5.

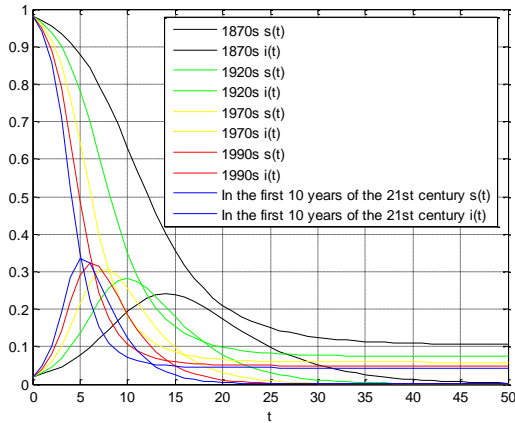


Figure. 5 the proportion of unknown persons  $s(t)$  and disseminators  $i(t)$  of five times trends over time

Due to a proportion of unknown person  $s(t)$  can fully describe the situation with the information communication trends over time, so this article only to proportion of unknown persons  $s(t)$  use the gray GM (1,1) prediction model for prediction. Use five times SIR data model of every day as the original data, to five times as the time line, use the gray GM (1,1) prediction model for prediction. Since the data is simulated to use its initial value of five times by changing the SIR model and actual data by a certain gap, so this article only gives a way to solve this problem. Using the gray GM (1,1) forecast model incorporates data five times per day, used on GM (1,1) model to forecast every day respectively, since the first day is the initial value of the data, five times is the same data, so to the next day, for example, obtained by today's information communication situation as shown in Table 5.

**Table 5. The next data dissemination of information of the five periods**

number	1	2	3	4	5
Unknown person	0.9687	0.963	0.9564	0.949	0.9404

Step 1 Ratio Test level

Establish the number of unknown proportion of five periods of time-series data as follows,

$$x_1^{(0)} = (x_1^{(0)}(1), x_1^{(0)}(2), \dots, x_1^{(0)}(5)) \\ = (0.9687, 0.9630, 0.9564, 0.9490, 0.9404)$$

Level ratio

$$\sigma^{(0)}(k) \\ \sigma^{(0)}(k) = x^{(0)}(k-1) / x^{(0)}(k) \\ \sigma^{(0)} = (\sigma^{(0)}(2), \sigma^{(0)}(3), \dots, \sigma^{(0)}(5)) \\ = (1.0059 \quad 1.0069 \quad 1.0078 \quad 1.0091)$$

judge Level ratio, since all

$$\sigma^{(0)}(k) \in [1.0059 \quad 1.0091], \quad k = 2, 3, \dots, 5$$

Therefore, you can use  $x^{(0)}$  as satisfactory GM (1,1) Model

Step2  $x_1^{(0)}$  accumulate the original data, namely

$$x_1^{(1)} = (0.9687 \quad 1.9317 \quad 2.8881 \quad 3.8371 \quad 4.7775)$$

Step3 Construction of a data series  $B$  and Vector Data  $Y$ .

$$B = \begin{pmatrix} -\frac{1}{2}(x_1^{(1)}(1) + x_1^{(1)}(2)) & 1 \\ -\frac{1}{2}(x_1^{(1)}(2) + x_1^{(1)}(3)) & 1 \\ \dots & \dots \\ -\frac{1}{2}(x_1^{(1)}(4) + x_1^{(1)}(5)) & 1 \end{pmatrix}$$

$$Y_n = (x_1^{(0)}(2), x_1^{(0)}(3), \dots, x_1^{(0)}(5))$$

Step4 Compute  $\hat{a}$

$$\hat{a} = (B^T B)^{-1} B^T Y = \begin{pmatrix} 0.0079 \\ 0.9750 \end{pmatrix} = (a, u)^T$$

Step5 Found Model

$$\frac{dx_1^{(1)}}{dt} + 0.0079 x_1^{(1)} = 0.9750$$

take  $x_1^{(1)}(0) = x_1^{(0)}(1) = 0.9687$  obtaining the solution (referred to as time response function) to

$$\hat{x}_1^{(1)}(k+1) = (x_1^{(0)}(1) - \frac{u}{a})e^{-ak} + \frac{u}{a} = 123.504 - 122.536e^{-0.789409-2k}$$

Step6 Seeking to generate the number of column values

$\hat{x}_1^{(1)}(k+1)$  and restore the value of the model.

Make  $k=1, 2, \dots, 4$ , from the function of the response time above can be considered  $\hat{x}^{(1)}$ , and take

$$\hat{x}^{(1)}(1) = \hat{x}^{(0)}(1) = x^{(0)}(1) = 0.9687$$

by  $\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1)$ , take  $k=2, 3, \dots, 5$ , get

$$\hat{x}^{(0)} = (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(5)) \\ = (0.9687 \quad 0.9635 \quad 0.9559 \quad 0.9484 \quad 0.9409)$$

4.2.2.3 Result analysis

Based on information dissemination of the model the original value and the model value of the five times, as well as residuals of the model, relative error and the average of accuracy in Table 6.

**Table 6 Prediction and forecasting accuracy of the model**

No.	Original value	Model values	Residuals	Relative error	Residual level
1	0.9687	0.9687			
2	0.9630	0.9635	-0.0004997	0.0519%	0.0020
3	0.9564	0.9559	0.0004763	0.0498%	0.0010
4	0.9490	0.9484	0.0005927	0.0625%	0.0001

Table 6 we can get, the accuracy of the model are higher, so it can be used to predict today's information communication situation

$$\hat{x}^{(0)}(6) = \hat{x}^{(1)}(6) - \hat{x}^{(1)}(5) = 0.9389$$

4.2.2.4 Inspection of GM (1,1) Model

Since our analysis of 2050 later in the relationship between communication and the ability to use the network in the case of today's information communication prediction data GM (1,1) Model. To illustrate the effectiveness of the prediction and assessment of the results, we will test to GM (1,1) model,



testing accuracy of GM(1,1) model , we will use the method of residual test .

The definition  $k$  point (or time)of the original time series  $x^{(0)}$  is  $x^{(0)}(k)$  .form  $x^{(0)}$  obtained the calculation of gray model is  $\hat{x}^{(0)}(k)$  called  $q(k) = x^{(0)}(k) - \hat{x}^{(0)}(k)$  is residuals of  $k$  point (or time) ; Relative error  $\varepsilon(k)$  and average relative error  $\varepsilon(avg)$  are defined as follows:

$$\varepsilon(k) = q(k) / x^{(0)}(k) * 100\%$$

$$\varepsilon(avg) = 1 / (n - 1) \sum_{k=2}^n |\varepsilon(k)|$$

For  $\varepsilon(k)$  , the general requirements  $\varepsilon(k) < 20\%$  , preferably  $\varepsilon(k) < 10\%$  . Solving the test results of GM (1,1) model using MATLAB , we get  $\varepsilon(avg) = 1.89\% < 10\%$  .It showed that use GM (1,1) model to predict today's information communication situation is entirely possible.

Similarly, to predict on the second and third information communication skills, can get communication of information available the trend over time today , as shown in Figure 6.

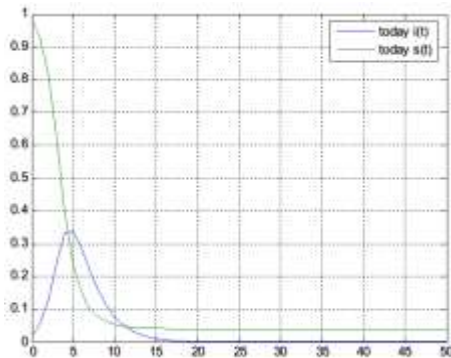


Figure 6 Map information communication trends today forecast

## 5. MODEL EVALUATION

### 5.1 Analysis of the advantages of the model

In this paper, conduct a comprehensive study to the flow of information ,the information flow has similarities with the spread of infectious diseases. So use the SIR model to model of the information flow, according to data obtained SIR model based on least squares polynomial fitting, then qualitative analysis to the SIR model, and the use of reliable data be verified to ensure that the flow of information on the SIR model established with a certain reliability.

Since there is no problem to get the information in the data flow to five times, using a modification of SIR model the initial value data obtained as the data flow of information five times, and then use the gray GM (1,1) model to predict for today and the flow of information in 2050.Although the initial data is simulated, there is no certain practical meaning, this model gives an effective way to solve such problems.

Prediction model network environment can be a good characterization of the news media. On this basis, this paper qualitative analyzes how in today's Internet world, so that the public interest and views are changed by information networks.

Topology or intensity form of messages or sources , information networks and how they are used to disseminate information and influence public opinion, studies on the links between social network information and to make a reference.

### 5.2 Analysis disadvantage of the model

Since news events related to network news, forums, blog, microblogging and other Internet information resources, information capacity, high redundancy and susceptible to noise interference. To collect comprehensive information on these and filtering, on the one hand a huge workload and high labor costs, On the other hand due to the passage of time, some of the information on the Internet will disappear, to be covered by the new information, Therefore, quantitative study of this article only collect and use information representative corpus as samples. Although, in the foregoing chapter we have made under the existing study based on the results of the theory and use of the public emergency network news also from a public opinion which reflects the evolution of the development and changes of the network news coverage. However, Insufficient sample of the present study was to analyze the qualitative part brings some limitations.

Since public opinion guidance and management is a very practical work, specific to the implementation level. We need to deal with a large number of complex issues and deal with complicated things, due to the lack of implementation experience, guidance and management research in the public emergency network public opinion aspects, This article does not find significant innovations, suggestions and measures proposed are mainly to further enrich the existing research-based. . In the future, we prefer to integrate the optimization with some traditional data mining approaches [3], [4], [5] to find more effective ways for optimization.

## 6. REFERENCES

- [1] Changjie Chen, YiMing Wei: Based Information Sharing for Sustainable Development Evaluation and Optimization [J](Systems Engineering:2011)
- [2] Marshall T. Poe, A History of Communications: Media and Society from the Evolution of Speech to the Internet (New York: Cambridge, 2011)
- [3] Ting Lu, Qi Gao, Xudong Wang, Cong Liu, "Modified Page Rank Model in Investigation of Criminal Gang", International Journal of Science and Engineering Applications, 4(3): 100-104, 2015.
- [4] Baohua Liu, Xudong Wang, Qi Gao, Cong Liu, "A K-Means based Model towards Ebola Virus Prorogation Prediction", International Journal of Science and Engineering Applications, 4(3): 95-99, 2015.
- [5] Libao Zhang, Faming Lu, An Liu, Pingping Guo and Cong Liu\*, "Application of K-Means Clustering Algorithm for Classification of NBA Guards", International Journal of Science and Engineering Applications, 5(1), pp. 1-6, 2016.

# A Review on Strategies of Word Sense Disambiguation

Roshan G. Belsare  
PRMIT&R, Badnera, Amravati

Prof. Shrikant P. Akarte  
PRMIT&R, Badnera, Amravati

---

**Abstract:** Word sense disambiguation is an important and challenging task in natural language processing. Its goal is to find the correct sense in which a word occurs in a sentence or a query when it can have multiple meanings. It is used in various applications of NLP like machine learning, text summarization, information retrieval etc. In this paper, we made a survey of supervised, unsupervised, knowledge based and corpus based approaches of word sense disambiguation. In this paper, study of various word sense disambiguation strategies has been done.

**Keywords:** WordSense Disambiguation, supervised, unsupervised, knowledge based

---

## 1. INTRODUCTION

Most natural languages contain *polysemous* words that is, the words that have the same spelling but have different possible meanings or *senses*. For example, in English the word *bark* can refer to the sound made by a dog or the covering of a tree. Humans are naturally good at identifying which sense of the word is used in a particular sentence. For instance, take the sentence *The old lady got scared hearing the bark of the dog*, we immediately come to know that *bark* is used to refer to the sound made by the dog whereas given the line *the dog scratched its back on the bark of the tree* we know that bark here means covering of the tree.

However, this knowledge of human beings is based vast experience of the world as well as languages which lacks in computer programs, hence determining the correct sense applied according to the context is a difficult task. The process of differentiating between the different meanings of a polysemous word and assigning the correct meaning or sense to it is Word Sense Disambiguation. WSD is considered as AI complete problem [5].

Difficulty in WSD is due to two aspects. First, Dictionary based glosses tend to be ambiguous. Different lexicographers may tag different senses to the same instance. Second, WSD involves much world knowledge or common sense, which is difficult to verbalize in dictionaries [6].

The task description of WSD can be formulated as a method of assigning the appropriate sense to all or some words in the text  $T$  where  $T$  is a sequence of words  $(w_0, w_1, \dots, w_{n-1})$  to find the mapping  $M$  from words to senses such that  $M(k) \subseteq \text{Senses } J(w_k)$  where  $M(k)$  is the subset of senses of  $w_k$  which are appropriate in the text  $T$  and  $J(w_k)$  is the set of senses in dictionary  $J$  for word  $w_k$ .

The mapping  $M$  can assign more than one sense to  $w_k$  belonging to  $T$  but eventually the most appropriate sense is selected. Thus WSD is a classification task where word senses are the classes and the classification method classifies each occurrence of the word to more than one class based on external knowledge sources and context.

The paper has been further divided into six sections. In section II a brief discussion of history of the research done in WSD has been given. Section III gives a brief discussion of

the knowledge based approaches. In section IV supervised disambiguation approach has been highlighted followed by unsupervised disambiguation approach in section V. Conclusion has been discussed in section VI.

## 2. LITERATURE SURVEY

In 1940s WSD was first formulated as separate computational task during the early days of machine translation. This makes it one of the oldest problems in computational linguistics. In 1949, Warren Weaver first introduced the concept in computational context.

Till 1970s WSD was a subtask of semantic interpretation systems which were developed within the field of artificial intelligence. However, since WSD systems were at the time largely rule-based and hand-coded they were prone to a knowledge acquisition bottleneck. In the 1990s, the statistical revolution swept through computational linguistics, and WSD became a paradigm problem on which to apply supervised machine learning techniques. Since then, supervised techniques have reached a plateau in accuracy, and so attention has shifted to coarser-grained senses, domain adaptation, semi-supervised and unsupervised corpus-based systems, combinations of different methods, and the return of knowledge-based systems via graph-based methods. Still, supervised systems continue to perform best.

## 3. KNOWLEDGE BASED APPROACH

The idea behind the knowledge based approach is to make extensive use of knowledge sources to decide upon the senses of words in a particular context. It was found that although alternate supervised approaches were more efficient than knowledge based approaches but their advantages also covered a wide range. Collocations, thesauri, dictionaries etc are the most commonly used resources in this approach. Initially knowledge based approaches started in limited domains in 1979 and 1980 [8]. There are three Knowledge based approaches which are discussed as follows:

### 3.1 Lesk Algorithm

M. Lesk proposed a approach to determine the overlap between words in the sense definitions of ambiguous words and the definitions of context words surrounding these ambiguous words in a given text. The biggest drawback of this algorithm is that dictionary definitions are often very short (as lexico and do not have enough



words for this algorithm to work well. A modification has been proposed by Banerjee and Pedersen [9] that deals with this problem by adapting this algorithm to the semantically organized lexical database called WordNet. Besides storing words and their meaning like a normal dictionary, WordNet also "connects" related words together. They overcome the problem of short definitions by looking for common words not only between the definitions of the words being disambiguated, but also between the definitions of words that are closely related to them in WordNet. Their algorithm achieves an 83% improvement in accuracy over the standard Lesk algorithm, and that it compares favorably with other systems evaluated on the same data..

### 3.2 WSD using conceptual density

The conceptual density is the measure of how the concept that the word represents is related to the concept of the words in its context. Conceptual density is related to conceptual distance inversely. The conceptual distance is determined from the WordNet.

### 3.3 Walker's Algorithm

Walker proposed a simple algorithm by incorporating subject codes. His algorithm is based on the assumption that the subject codes assigned to a word reflects the sense of the word. If a word has more than one subject code then it will have more than one sense. For example: Longman's Dictionary of Contemporary English includes subject code EC (economic) for the "financial" sense of "bank". This subject code helps us in knowing that "deposit" is related to the "financial" sense of bank [4] [10].

His algorithm is based on the assumption that the subject codes assigned to a word reflects the sense of the word. If a word has more than one subject code then it will have more than one sense. For example: Longman's Dictionary of Contemporary English includes subject code EC (economic) for the "financial" sense of "bank". This subject code helps us in knowing that "deposit" is related to the "financial" sense of bank [4] [10].

## 4. UNSUPERVISED APPROACH

Unsupervised methods of WSD eliminate the need for sense tagged training data and therefore, they overcome the knowledge acquisition bottleneck [1].

Strictly speaking, using a completely unsupervised sense disambiguation task, we can only discriminate word senses. That is, we can group together instances of a word used in different senses without knowing what those senses are.

However, Yarowsky [3] proposed an unsupervised algorithm that can accurately disambiguate word senses in a large completely untagged corpus. He exploited two powerful properties of human language in an iterative bootstrapping setup to avoid the need of manually tagged training data (adapted from Yarowsky 1995)[4]:

1. One sense per discourse: The sense of a target word is highly consistent within any given document or discourse.
2. One sense per collocation: Nearby words provide strong and consistent clues to the sense of a target

word, conditional on relative distance, order and syntactic relationship.

This approach has two types of distributional approaches; first one is monolingual corpora and other one is translation equivalence based on parallel corpora. And these techniques are further categorized into two types; type-based and token-based approach. The type-based approach disambiguates by clustering instances of a target word and token-based approach disambiguates by clustering context of a target word.

#### A. Context Clustering

In Context Clustering method [2], first context vectors are created and then they are grouped into clusters to identify the meaning of the word. This method uses vector space as word space and its dimensions are words only. Also in this method, a word which is in a corpus will be denoted as vector and the no of times it occurs will be counted within its context. After that, co-occurrence matrix is created and similarity measures are applied. Then discrimination is performed using any clustering technique.

#### B. Word Clustering

This technique is similar to context clustering in terms of finding sense but it clusters those words which are semantically identical. For clustering, this approach uses Lin's method. It checks identical words which are similar to target word. And similarity among those words is calculated from the features they are sharing. This can be obtained from the corpus. As words are similar they share same kind of dependency in corpus. After that, clustering algorithm is applied to discrimination among senses. If a list of words is taken, first the similarity among them is found and then those words are ordered according to that similarity and a similarity tree is created. At the first stage, only one node is there and for each word available in the list, iteration is applied to add the most similar word to the initial node in the tree. Finally, pruning is applied to the tree. As a result, it generates sub-trees. The sub-tree for which the root is the initial word that we have taken to find sense, gives the senses of that word. Another method to this approach is clustering by committee. As mentioned earlier, the word clustering is approach is clustering by committee. As mentioned earlier, the word clustering is a kind of context clustering, this clustering by committee follows similar step, first the similarity matrix is created, so that, matrix contains pair-wise similar information about the words. And in the next step, average-link clustering is applied to the words. The discrimination among words is performed using the similarity of centroids. For each committee, one centroid exists. So, according to the similarity of the centroid, the target word gives the respective committee. In the next step, features between the committee and the word are removed from the original word set, so in next iteration, identification of senses for same word which are less frequent, is allowed.

#### C. Co-occurrence Graph

This method creates co-occurrence graph with vertex V and edge E, where V represents the words in text and E is added if the words co-occur in the relation according to syntax in the same paragraph or text. For a given target word, first, the graph is created and the adjacency matrix for the graph is created. After that, the Markov clustering method is applied to find the meaning of the word.

Each edge of graph is assigned a weight which is the co-occurring frequency of those words.

Weight for edge {m,n} is given by the formula:

$$w_{mn} = 1 - \max\{P(w_m | w_n), P(w_n | w_m)\}$$

Where  $P(w_m | w_n)$  is the  $\text{freq}_{mn} / \text{freq}_n$  where  $\text{freq}_{mn}$  is the co-occurrence frequency of words  $w_m$  and  $w_n$ ,  $\text{freq}_n$  is the occurrence frequency of  $w_n$ . Word with high frequency is assigned the weight 0, and the words which are rarely co-occurring, assigned the weight 1. Edges, whose weights exceed certain threshold, are omitted. Then an iterative algorithm is applied to graph and the node having highest relative degree, is selected as hub. Algorithm comes to an end, when frequency of a word to its hub reaches to below threshold. At last, whole hub is denoted as sense of the given target word. The hubs of the target word which have zero weight are linked and the minimum spanning tree is created from the graph. This spanning tree is used to disambiguate the actual sense of the target word.

## 5. SUPERVISED APPROACH

Approaches relying on sense tagged corpora for disambiguation are known as supervised. WSD approaches. They yield very high accuracy in the domain of the training corpus. But this accuracy comes at the cost of sense tagged corpora which is a costly resource in terms of the time and the manual efforts involved. Creating such corpora for all languages in all domains will be impracticable. Hence these approaches cannot be easily ported to different languages or domains. Some good supervised approaches are mentioned below [8].

### A. Decision Tree

A decision tree [11-12] is used to denote classification rules in a tree structure that it recursively divides the training data set. Internal node of a decision tree denotes a test which is going to be applied on a feature value and each branch denotes an output of the test. When a leaf node is reached, the sense of the word is represented (if possible). For example, The noun sense of the ambiguous word “bank” is classified in the sentence, “I will be at the bank of Narmada River in the afternoon”.

### B. Neural Networks

In the Neural Network based computational model, artificial neurons are used for data processing using connectionist approach. The input includes the input features and the target output and goal is to partition the training context into non-overlapping sets. The training dataset is divided into sets which are non-overlapping based on desired responses. When the network encounters new input pairs the weights are adjusted so that the output unit giving the target output has the larger activation. The network can have weights both positive and negative corresponding to correct or wrong sense choice. Neural networks can be used to represent words as nodes and these words will activate the ideas to which they are semantically related. The inputs are propagated from the input layer to the output layer through the all intermediate layers. The input can easily be propagated through the network and manipulated to arrive at an output. It is difficult to compute a clear output from a network where the connections are spread in all directions and form loops.

## C. Naïve Bayes

Naive Bayes classifier is the classifier based on Bayes theorem and assumes that every feature is class conditionally independent of every other feature. This approach classifies text documents using two parameters: the conditional probability of each sense ( $S_i$ ) of a word ( $w$ ) and the features ( $f_j$ ) in the context.

WSD is very tough problem and needs large number of lexical and knowledge resources like sense tagged corpora, machine readable dictionaries *etc.* It is evident that use of such resources improves the performance of WSD. Hence one might think that, if such resources are available, and then why not use them? Or why not spend sufficient time in creating high quality resources and perform great in terms of accuracy. The main reason is that, even if we have all possible resources to build a great supervised approach, it cannot be ported to other language easily. The resources have to be replicated for all possible languages. Another disadvantage of using the supervised approaches is, by using fixed sense repositories; we constrain our self to the fixed number of senses present in that repository. We cannot discover new senses of words, which are not present in the sense repository. Hence only considering the accuracy of the approach is not a good idea, but considering its versatility and portability to other languages and domains is also equally important. This is the reason we see many unsupervised approaches being tried by many researchers in WSD [7].

## 6. CONCLUSION

WSD is a very complex task in Natural language processing as it has to deal with complexities found in a language. In this paper we have put forwarded a survey of comparison of different approaches available in word sense disambiguation with primarily focusing on the knowledge based, supervised and unsupervised approaches. We concluded that supervised approach is found to perform better but one of its disadvantage is the requirement of a large corpora without which training is impossible which can be overcome in unsupervised approach as it does not rely on any such large scale resource for the disambiguation. Knowledge based approach on the other hand makes use of knowledge sources to decide upon the senses of words in a particular context provided machine readable knowledge base is available to apply.

## 7. REFERENCES

- [1] Gale, W. A., Church, K., and Yarowsky, D. 1992b. A method for disambiguating word senses in a corpus. *Comput Human.* 26, 415–439.
- [2] Niu, C., Li, W., Srihari, R. K., Li, H., Crist, L., (2004) “Context Clustering for Word Sense Disambiguation Based on Modeling Pairwise Context Similarities”, *SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, Barcelona, Spain, July 2004.
- [3] Yarowsky, D., 1995, ‘Unsupervised word sense disambiguation rivaling supervised methods,’ *Proceedings of the 33<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA, pp. 189-96.
- [4] U. S. Tiwary, Tanveer Siddiqui, 2008, *Natural Language Processing and Information Retrieval*.
- [5] Samit Kumar, Neetu Sharma, Dr. S. Niranjana, “Word Sense Disambiguation Using Association Rules: A Survey”,

International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 2, 2012

[6] Veronis, J. 2000. Sense Tagging: Don't Look for the Meaning But for the Use, *Workshop on Computational Lexicography and Multimedia Dictionaries*, 1-9. Patras, Greece.

[7]Alok Ranjan Pal, Diganta Saha, 2015, Word Sense Disambiguation: A Survey.

[8] J. Sreedhar, S. Viswanadha Raju, A. Vinaya Babu, Amjan Shaik, P. Pavan Kumar, “Word Sense Disambiguation: An Empirical Survey”, *International Journal of Soft Computing and Engineering (IJSCE)*, Volume-2, Issue-2, May,2012.

[9]Satanjeev Banerjee, Ted Pedersen, 2002, An Adapted Lesk Algorithm for Word Sense Disambiguation using WordNet.

[10]Walker, Donald, 1987, ‘Knowledge resource tools for accessing large text files,’ *Machine Translation: Theoretical and Methodological Issues*, Sergei Nirenberg(Eds.), *Machine Translatin of Languages*, John Wiley & Sons, New York, pp.

[11] Singh, R. L., Ghosh, K. , Nongmeikapam, K. and Bandyopadhyay, S.,(2014) “A DECISION TREE BASED WORD SENSE DISAMBIGUATION SYSTEM IN MANIPURI LANGUAGE”, *Advanced Computing: An International Journal (ACIJ)*, Vol.5, No.4, July 2014, pp 17-22.

[12]<http://www.d.umn.edu/~tpederse/Pubs/naacl01.pdf> date: 14/05/2015.

[13]Devendra Singh Chaplot, 2014, Literature Survey on Unsupervised Word Sense Disambiguation.

[14] Pranjal Protim Borah, Gitimoni Talukdar, Arup Baruah, , 2014, Approaches for Word Sense Disambiguation : A Survey, *International Journal of Recent Technology and Engineering (IJRTE)*,ISSN: 2277-3878, Volume-3, Issue-1, March 2014

# A Case Study of Volkswagen Unethical Practice in Diesel Emission Test

Nazanin Mansouri  
HELP University  
Kuala Lumpur, Malaysia

**Abstract:** By end of 2015, automotive industry was shocked by an unethical action of Volkswagen which is one of the noticeable market players amongst automakers. Since October 2015 this case has been under investigation by various authorities. The investigators have managed to find out significant factors leading to this emission scandal. Volkswagen deceptive scandal has resulted in a series of overwhelming devastating consequences on direct and indirect stakeholders of Volkswagen. Since this case is still ongoing, the judge has not succeeded to finalize the case and deliver Volkswagen penalties.

In this case, the participants and primary reasons of this rigging in diesel emission test are identified in line with previous studies. Moreover, the critical impacts of such deception on Volkswagen stakeholders are explained in details. The proposals which are given to the judge to consider an appropriate deal to Volkswagen are discussed in details. Last but not least, in order to prevent such scandal to take place in the future, recommendations are provided.

**Keywords:** Volkswagen emission test, Volkswagen deception, EPA standard, Environmental agency

## 1. INTRODUCTION

In the recent decades conservation and sustainable environment are of great concern to the nations. The natural phenomenon such as globalization and also environment pollutions such as emission have been considered by a great number of environmental protection agencies and united nation agencies. A noticeable attention to conservation and saving green have led Environmental Protection Agency (EPA) tighten emission control because of harmful and mortal effects of nitrogen oxide which is a pollutant found in car's exhaust (Klier, & Linn, 2016). Therefore, since 1970s EPA has announced continuously more restrictions on standard of emission for light-duty vehicles entailing small pickup trucks, automobiles, and sport-utility cars. The most stringent requirements for emission standards were for vehicle models of year 2004 (Klier, & Linn, 2016). Not only EPA but also federal agency created significant diminution (94%) in the amount of emitted nitrogen oxide by vehicles tailpipe from 1.25 to 0.07 grams per mile (Klier, & Linn, 2016). Emitted nitrogen oxide endangers human lives and triggers disease such as asthma, respiratory, cardiovascular, bronchitis, and pre-mature death.

The new emission standard posed immense hardship to automakers manufacturing fuel-efficient diesel cars to the United States automobile market. One of the market players in automobile industry is Volkswagen attempting to crack the United States diesel market; as a result, Volkswagen became a substantial seller in automaker market. Volkswagen competitors namely Honda, Hyundai, Mazda, and Nissan found new emission standards significantly challenging; therefore, they made decision to scrap their tactics (Davenport, & Hakim, 2016).

Surprisingly, in the years 2015, Volkswagen was announced as "diesel dupe" in virtue of rigging emission test to make diesel vehicles seem emitting less pollution than what they really emit (Clothier, 2015). In September 2015 it was reported by EPA that in an ample number of Volkswagen vehicles, sold in worldwide, a defeat device or software was embedded in diesel engine with the purpose of changing vehicle performance to improve required result. Volkswagen aimed at pretending that its vehicles follow emission standards; therefore, conducted emission test in the lab instead

of on the roads (Le Page, 2015). Volkswagen cars were programmed to detect the situation where cars with TDI diesel engines experience emission test and then take information from brakes, accelerator, and steering. Subsequently, the program made slight changes to engine setting with the purpose of diminishing nitrogen oxide level emitted by Volkswagen diesel cars. The actual result of Volkswagen emission test on road was thirty-five (35) times more than cheated result in the lab (Le Page, 2015). Volkswagen emission test scandal has created a dramatic consequences having impact on ample number of authorities.

## 2. FINDINGS

3. There have been animated debates surrounding the Volkswagen diesel scandal to find out primary reasons that what factors paved the ground for such scandal to take place and also who was involved in this scandal. This scandal has been reported in a great number of media and news.

### 2.1. Participants in Volkswagen Emission Test Scandal

Firstly, it was reported that nine (9) managers are suspended for being involved in the deception. It is stated by Volkswagen chairman that one of the reason suspended group took this unethical scandal is because of the mindset in some departments of the company which fails to comply the rules (Goodman, McGrath, & Leah, 2015).

In October 2015, it was stated that the group of employees participated in diesel scandal is small. CEO of Volkswagen announced to the United States lawmakers that for whatever reasons a small group of software engineers embedded the device in cars. The CEO mentioned that he is not aware of exact number of engineers and also this scandal was not corporate decision (Boston, Varnholt, & Sloat, 2015).

Later, a law firm Jones Day, which is the external investigator of Volkswagen, conducted internal investigation and later revealed that fifty (50) members of staff mostly in Wolfsburg confessed that they were completely aware of emission scandal activities.

Later, news learned that although some Volkswagen engineers and technicians informed their supervisors about emission

rigging activities in the year 2011, supervisors ignored the alert (Boston, Varnholt, & Sloat, 2015).

In November 2015, the number of participated managers, technicians, and engineers in the emission test scandal to operate the defeat device is dramatic. There have been various levers in this scandal. It is described that one individual is able to write the codes of defeat device program; on the other hand implementing this program to function with engine and other elements requires more workforces to be involved in (Goodman, McGrath, & Leah, 2015). Moreover, in accordance with Bode, Volkswagen junior positions were not included in the cheating test, whilst finally it was declared that a great number of managerial and non-managerial the rank and file workforce were involved in diesel emission scandal. The suspended managerial group consists of nine (9) managers from supervisory board, quality control and assurance, engine designers, plant managers, and managers of sub-companies such as Porsche and Audi. EPA has added that approximately hundred (100) million lines of codes are included in each car, whilst in the case of Volkswagen; the defeat device was embedded amongst millions of codes which are definitely far challenging to detect. In addition, there is a possibility that a third-party programmer has written the codes of cheating program, Ricker added (Boston, Varnholt, & Sloat, 2015).

## 2.2. Reasons for Rigging Emission Test

Martin Winterkorn Volkswagen CEO resigned on September and made announcement to decline that he was aware of diesel scandal. It is believed greatly by a great number of individuals that this unethical scandal occurred because of technology shortcoming prompted Volkswagen to cheat in emission test (Boston, Varnholt, & Sloat, 2015). On December Hans-Dieter Pötsch chairman of Volkswagen revealed to the public that a group of Volkswagen engineers made decision to rig diesel emission test in 2005 when the united States EPA imposed the toughest emission standards to automobile industry. The significant reason forced them to take this decision at that time was in virtue of technology shortcoming (Boston, Varnholt, & Sloat, 2015). The engineers were unable to meet the United States emission standards through Volkswagen given time and allocated budget. After specific time span, Volkswagen engineers managed to find a solution to face emission standards, yet they preferred to continue rigging rather than implementing the method. In accordance with Volkswagen chairman public admission, the unethical scandal is not one-off mistake, but it is a whole chain of mistakes taking place without any disruption (Rising & Sopke, 2015).

One of the other significant reasons fueling Volkswagen deception is unique corporate culture of the company. The culture of Volkswagen is compliance based which employees are obliged to comply with the rules (Goodman, McGrath, & Leah, 2015). This culture has led to an atmosphere where workforces carry out their tasks under a critically centralized structure. In accordance with this culture, the demand and expectations of the company should be fulfilled regardless of how employees are able to perform the tasks (Goodman, McGrath, & Leah 2015). According to director of the Center for Automotive Research Ferdinand Dudenhöffe, Volkswagen cultural practices are far different from other automobile manufacturers. The company is autocratic rather than being democratic. The main focus of the company is on roots and Wolfsburg; therefore, there is a noticeable dearth of global thinking. Based on Dudenhöffer, chief positions of Volkswagen might not directly enforced workforces to install defeat device in engines however working environment of

Volkswagen is famous for avoiding dissent and discussion. In Volkswagen human resource practices pave the ground for management positions to tell employees to consider the task again, and if you are not capable to find a solution or to perform your tasks, there are other candidates who are efficient to perform your job. Thus employees of Volkswagen find themselves in a situation where if maintaining job is the concern, there is no choice (Sherk, 2014).

In accordance with one of Volkswagen executive employee, rewarding system of Volkswagen is another reason enforcing employees to prefer keeping quiet about rigging in emission test for a long time even if they preferred to reveal. The bonus system of Volkswagen applies to all workforces from lower positions such as assembly line to higher positions such as chief executives (Elson, Ferrere, & Goossen, 2015). In Volkswagen rewarding and bonus system, not only employees based on their performance receive bonus but also teams based on their performance and productivity receive bonus. The financial incentive of Volkswagen rewarding system leads employees not to come out with dissenting opinions. In accordance with one of Volkswagen employees, bonus is in commensurate with the amount of remuneration that a staff is paid. In other words, when a Volkswagen staff takes up a greater position, the amount of bonus that staff receives is higher than an assembly line staff that is on lower remuneration (Elson, Ferrere, & Goossen, 2015).

When Volkswagen cheating in diesel emission test was admitted publically, Bernd Osterloh who is the labor leader of Volkswagen declared that culture and approaches of company is not ethical and appropriate; therefore, some value-based changes should be adapted in the culture of company to allow employees to communicate with higher levels openly regarding any matter within the company, and also to enable employees to share their dissent opinion with their supervisors or chief positions (Sherk, 2014).

Although some people might stick to the idea that Volkswagen scandal took place in virtue of profit maximization, this is in face a misconception. According to the revealed information, it is evident that the primary roots of this deception are fueled from Volkswagen culture and approaches.

## 3. ANALYSIS

Volkswagen scandal has led to a devastating series of critical consequences to its direct and indirect stakeholders. This emission scandal is still under investigation by various authorities with the purpose of identifying actual impacts and coming out with a suitable solutions to finalize the case.

### 3.1 Impact of Volkswagen Diesel Scandal

In the year 2014, Volkswagen was the world's second largest automaker in automobile industry after Toyota Motor Corporation. In the year 2015, Volkswagen admitted rigging in diesel emission test which caused the company to suffer from huge amount of cost burden. Volkswagen has brought in three (3) public relations firms based in United States, Britain, and Germany to assist the company to cope with the crisis. Moreover, Volkswagen has employed the former communications of BMW as a consultant to work 60 hours a week with salary of \$22,000 per month (Hakim, 2016). Moreover, since this case includes various countries over the world, Volkswagen is required to deal with different international regulations. The primary consequences of this unethical deceptive scandal are listed as follow (Hakim, 2016).

#### 3.1.1 Threatening People's Health



By end of 2016, Volkswagen cars with defeat device will have produced additional toxic pollution to directly trigger premature death of roughly sixty (60) individuals merely in the United States. From the year 2008 to 2015, 428,000 Volkswagen and Audi diesel cars pumped out nitrogen oxide forty (40) times more than it was allowed by the Clean Air Act. It is estimated by the researchers that with six (6) years Volkswagen and Audi diesel produced an excess of 36.7 million kg nitrogen oxide to the environment (Selin, 2015b). Nitrogen oxide is a primary element of particulate and smog matter which paves the ground for various disease namely heart disease, premature death, bronchitis, and respiratory and cardiovascular disease. Researchers have estimated that significant impacts of nitrogen oxide produced by Volkswagen cars endanger 60 human lives from 10 to 20 years prematurely (Selin, 2015a).

It is noted that excess of pollution from Volkswagen vehicles participated directly in thirty-one (31) and thirty-four (34) chronic bronchitis and admission of respiratory and heart cases respectively in the United States. Additional pollution to the environment will result in 120,000 minor restricted activity day and approximately 210,000 days of less respiratory signs. The sickness of people over six (6) years from 2008 to 2015 will cost United States \$450 million (Kalaugher, 2015).

If Volkswagen declines to recall vehicles with defeat device, from 2015 onwards 140 premature deaths will take place. In addition, health cost of \$840 million will be caused by the Volkswagen diesel cars (Chue, 2015).

In addition, excess of nitrogen oxide to the natural environment by Volkswagen diesel cars results in acid rains (The Editorial Board, 2016). The acid rains not only have crucial impact on human health but also have vital destruction on nature and natural resource.

### 3.1.2 Slump in Workforce's Bonus

When rigging in diesel emission test was revealed to the public, sales of Volkswagen was affected. Therefore, Volkswagen in order to cope with crisis, has announced that bonus of chief management will be reduced substantially (Moulson, 2016). Volkswagen said a statement that various models which establish fair and rational solution for all participated parties are being deliberated. As a result, this leads to a considerable diminution of variable remuneration. The reduction in bonus will include management board and also a group of executive positions assisting CEO to operate the company's daily routine. It is said that the bonus of German equivalent of board of directors, which is the supervisory of Volkswagen, would not be reduced except Volkswagen chairman Hans Dieter Poetsch (Moulson, 2016).

### 3.1.3 Drop in Volkswagen Sale

Although Volkswagen diesel scandal has resulted in crucial impact on Volkswagen group brands namely Audi and Skoda, slump in sales of Volkswagen vehicles are far substantial than group brands. The bad reputation of Volkswagen has affected its customer loyalty; therefore, costumers switch from Volkswagen to its competitors which lead to a noticeable drop in sales (Kottasova, 2015). Since 2002, for the first time, in 2015 sales of Volkswagen plunged world-wide substantially in virtue of deceptive scandal. The following figure represents that Volkswagen's sales is continuously plunging in the United States (BBC, 2016; Sky UK, 2016; Waecsh, 2016).



Figure 3.1: Ongoing Drop in Volkswagen Sales  
Source: (Mittleman, 2016)

### 3.1.4 Plunge in Volkswagen Shares

Unethical practice of Volkswagen led to a dramatic slump in share value. Immediately once Volkswagen scandal was revealed, market showed reaction thus share value of the company dropped by one third. In other words, the emission scandal wiped billion dollars from Volkswagen value (Gomez, 2016). The following figure depicts that Volkswagen stocks price is continuously slumping dramatically in the United States.



Figure 3.2: Ongoing Drop in Volkswagen Stock Price  
Source: (Gomez, 2016)

### 3.1.5 Creating Hassel for Volkswagen Dealers

When deceptive emission took place, a package of specific programs with the purpose of assisting retailers to cope with Volkswagen rigging in diesel emission test was provided to Volkswagen dealers (Ausick, 2016). The program includes specific amount of money in form of sales bonus, incentives, or subsidy injected to dealership network struggling with lower sales and profit. The sale of Volkswagen diesel cars, which includes just above 20 percentages of total sales, was stopped. Therefore, dealers are still suffering from thin profit and sluggish sales (Beene, 2016).

On first of October Volkswagen of America with the purpose of relaxing crisis gave extra discretionary funds to Volkswagen dealers to use it in a way they wish. According to Automotive News, the amount of loan varied base on dealer's volume and reached to the highest amount of ten (10) thousand dollar. Volkswagen crisis grew significantly and affected the brand name, whilst dealers were not permitted to sell diesel vehicles. Thus dealers were concern whether the given fund is sufficient for survival or not. The CEO of Volkswagen America stated that further programs will be

considered for dealers with cash flow, but still specific date is not declared for programs (LaReau & Ryan, 2015).

A dealer of Volkswagen Steve Kalafer in New Jersey said that the given fund will be utilized to cover the store's operating costs and to boost marketing budget. Kalafer added that millions of dollars have been expended to strengthen this brand, but the scandal is an international deception; therefore, Volkswagen customers, employees, and investment of dealers have been affected. A great number of employees are concern about their job security and they are dissatisfied to encounter with customers suspending them whether they were aware of company's scandal or not (LaReau & Ryan, 2015).

Another Volkswagen dealer in east coast said that dealers of Volkswagen in the United States are not making money or merely breaking even. Moreover, the dealer adds that the given fund assists to break even or present thin profit.

The other dealer of Volkswagen Bill Wallace in Stuart said that overall customer traffic has reduced significantly and also it is critical to convert shoppers to sales. Therefore, the discretionary fund assists to him to close the deals (LaReau & Ryan, 2015).

General Manager of Volkswagen Tom Backer in New York said that some amount of the given fund will be utilized primarily to close deals with the owner of diesels cars who are unwilling to get into Volkswagen gasoline cars, whilst the rest of discretionary fund will be given to sales workforces with the purpose of improving their satisfaction and motivation (LaReau & Ryan, 2015).

### *3.2 Current Solutions for Finalizing Volkswagen Scandal*

On 21st April 2016 a judge in California United States declared that Volkswagen has offered a concrete plan which has been supported by United States government agencies and environmental protection agency to solve and finalize the case which Volkswagen cheated in emission test by installing a defeat device in car's engine to game emission standards of EPA in the United States. Almost end of 2015 Volkswagen admitted that in eleven (11) million cars globally the deceptive software has been installed and 600,000 of those cars are on United States roads. Therefore, agreement in principle will be applied to Volkswagen cars in the United States (Sorokanich, 2016). Judge Charles Breyer, U.S. District Court in San Francisco, has proposed a deal which gives options to owners of Volkswagen tow (2) liter diesel cars to fix their cars based on emission standards of the United States or sell back their cars to Volkswagen and cancel their loans (if the cars was purchased with loan). In accordance with the judge, the proposed deal includes considerable compensation for car owners of 2 liter diesels and also offers a fund for proper remediation hardship for any environmental harmful effects caused by Volkswagen cars exceeding pollutions to the environment. Still the details of compensation considered for Volkswagen car owners are not revealed (Goodman, et al., 2016a). Moreover, compensations for 3 liter diesels Volkswagen car owners are still underway. The deliberation of Volkswagen scandal is kept confidential. In accordance with the judge, the proposed agreement is approved and supported by California regulators, California attorney general's office, and the United States Department of Justice respectively. Lawyers stated that it is expected that the Federal Trade Commission (FTC) will accept the proposed deal. FTC recently sued Volkswagen for gaming in clean diesel test (Goodman, et al., 2016a).

The proposed agreement requires more approvals and comments from Volkswagen owners. It is said by the judge

that by July Volkswagen will be allowed to defend against a trial. The proposed agreement also consists of a section describing Volkswagen commitment to give further effort to create green car technology (Sorokanich, 2016).

Based on Kelly Blue Book, it costs Volkswagen more than seven (7) billion dollar to buy back the affected vehicles. Therefore, it is estimated that Volkswagen scandal might impose further cost to the company from approximately 6.7million euros to twofold or even more than the amount (Goodman, et al., 2016a).

Although Volkswagen and Porsche representatives, the Department of Justice, and others have given great efforts to work fourteen (14) hours a day to finalize an appropriate deal to solve Volkswagen deception, yet this case is not finalized yet. The judged announced that June 21st is the due date for all the involved parties to lodge their proposals before delivering the deal to public (Goodman, et al., 2016a). There is a proposal for Volkswagen emission scandal explained as follow.

It is estimated that a giant amount of money is about to be spent in seeking to penalize Volkswagen for its deception. It is proposed that instead of imposing additional financial burden to punish Volkswagen, there are other ways such as green way to panelize Volkswagen rigging in emission test (Schlanger, 2015). It has been recommended rather than squandering huge amount of money on penalties, it is more favorable to spend this amount of penalty to advance traditional technology. This is a great opportunity to bring out significant changes to automobile industry by requiring Volkswagen to concentrate on electric vehicles. In addition, there is no guaranty that Volkswagen will be able to repair the defeat device. On the other hand, the car owners might prevent to bring their car for fixing compromising performance (Yadigaroglu, 2015). It is rational that EPA releases Volkswagen accountability to recall existing vehicles with defeat device to fix them. Subsequently, EPA has a crucial role to enforce Volkswagen to speed up its roll-out of zero emission cars which emit zero pollutant to the environment. Therefore, zero-emission cars tighten opportunities for cheating in emission tests. Moreover, EPA benefits greatly from zero-emission cars since it is not required to create new complicated and expensive emission test for zero-emission cars. This type of cars has an ample of advantage not only to the EPA and automakers but also to the environment (Yadigaroglu, 2015). Over few years, Zero-emission cars significantly diminish emission which is ten (10) times greater than Volkswagen producing with deceptive device. In practical terms, increasing number of Volkswagen electronic vehicles on roads compensates the emission of Volkswagen cars with deceptive device. Besides, zero-emission cars privilege economy as well by increasing the needs of batteries. In order to fulfill the needs of battery in zero-emission cars, more battery manufacturers are required to be established over the world. High demand for battery manufacturing leads to an investment in lands and also creating jobs for locals in the United States and other regions. Advanced technology will be significantly reducing environment pollution, alleviating regulation of complicated emission standard, and also creating substantial domestic jobs. This is in facet an opportunity for Volkswagen to invest specific amount of money not only to compensate its scandal but also to strengthen its position. The proposed electronic cars have a dramatic role to repurpose the penalized money. According to the recommended proposal, it is evident that this is an ideal solution to finalize case of Volkswagen cheating in emission test (Yadigaroglu, 2015).

#### 4. CONCLUSION

In accordance with findings and analysis under previous sections, Volkswagen deceptive scandal is an overwhelming complicated case having created dramatic hassles for its direct and indirect stakeholders. Albeit Volkswagen cheating in diesel emission test was an unethical action resulting in series of disastrous consequences, the proposal of green punishment is a suitable solution to finalize this case. Electronic cars enable environment protection agencies to save cost and time for finding out new systems for controlling emission tests. Moreover, in this way, Volkswagen is not only penalized but also compensates its scandal appropriately. In conclusion, since this is an ongoing case, there is a chance to lodge the proposal of green punishment to judge to solve the issue in a win-win method.

#### 5. RECOMMENDATION

It is evident that the root of this unethical scandal goes back to business culture and structure of the company. Today's business practices are far different from the past practices. The compliance-based business ethics like Volkswagen approach declines to treat employees ethically and also employees face dilemma of loose job or take unethical action; therefore, the company obtains reverse result. The actions of employees have crucial role in success or failure of the company thus it is important to value employees and improve their moral to accomplish tasks ethically. In value-based practices hierarchy does not exist instead the employees work in teams to achieve expected results. Moreover, value-based business ethics promote democracy in the company which means executives have the freedom to voice their complaints to seniors or share their ideas within the company which might result in greater productivity.

In addition, it is obvious that emissions and pollutions are of significant concern to EPA. Thus it is undeniable that EPA new and strict emission standards imposed extreme pressure to automobile industry. EPA has an imperative role to alleviate the extreme pressure on automakers by providing and offering technology and research and development (R&D) aids. Case of Volkswagen reveals that stringent emission standards resulted in opposite expected outcome with a series of devastating events. Therefore, it is significant to plan for feasible emission standards and also provide technological and R&D support to automakers to prohibit such case to take place.

All in all, in order to prevent such scandal to occur in future, it is recommended to the companies to value their employees through value-based approach. Besides, EPA is recommended to provide supportive programs to automakers to ensure level of pollutants and emissions are under control.

#### REFERENCES

[1] Ausick, P. (2016, March 22). US Volkswagen Dealers Want VW to Pay, *24/7 Wall St.* Retrieved from <http://247wallst.com/autos/2016/03/22/us-volkswagen-dealers-want-vw-to-pay/>.  
[2] BBC, 2016, *Volkswagen branded sales drop for first time in 11 years*, online retrieved on 8<sup>th</sup> May 2016 from <http://www.bbc.com/news/business-35267164>.  
[3] Beene, R. (2016, March 31). VW dealers have a lot riding on the make meeting, *Automotive News.* Retrieved from <http://www.autonews.com/article/20160331/RETAIL06/304019986/vw-dealers-have-a-lot-riding-on-the-make-meeting>.

[4] Boston, W., Varnholt, H., & Sloat, S. (2015, December 10). Volkswagen Blames 'Chain of Mistakes' for Emissions Scandal, *The Wall Street Journal.* Retrieved from <http://www.wsj.com/articles/vw-shares-up-ahead-of-emissions-findings-1449740759>.

[5] Chu, J. (2015, October 28). Study: Volkswagen's emissions cheat to cause 60 premature deaths in U.S., *Massachusetts Institute of Technology.* Retrieved from <http://news.mit.edu/2015/volkswagen-emissions-cheat-cause-60-premature-deaths-1029>.

[6] Clothier, M. (2015, September 22). VW's U.S. Chief Apologizes for Diesel Dupe: 'We Screwed Up', *BloobBerg L.P.* Retrieved from <http://www.bloomberg.com/news/articles/2016-05-09/africa-s-would-be-switzerland-flaunts-economic-prowess-with-wef>.

[7] Davenport, C., & Hakim, D. (2016, January 4). U.S. Sues Volkswagen for Cheating on Emissions Tests, *The New York Times.* Retrieved from [http://www.nytimes.com/2016/01/05/business/vw-sued-justice-department-emissions-scandal.html?\\_r=0](http://www.nytimes.com/2016/01/05/business/vw-sued-justice-department-emissions-scandal.html?_r=0).

[8] Elson, C. M., Ferrere, C. K., & Goossen, N. J. (2015). The Bug At Volkswagen: Lessons in Co-Determination, Ownership, and Board Structure. *Journal of Applied Corporate Finance*, 27(4), 36-43.

[9] Gomez, A. (2016, January 4) Volkswagen (VLKAY) Stock Closes Down, U.S. Sales Drop 10%, *TheStreet.* Retrieved from <http://www.thestreet.com/story/13516097/1/volkswagen-vlkay-stock-closes-down-u-s-sales-drop-10.html>.

[10] Goodman, McGrath, & Leah, (2015, Dec. 15). Why Volkswagen Cheated; Intense Ambition and a Rigid Corporate Culture Created the Conditions for Lying at Germany's Biggest Company, *Newsweek.* Retrieved from <http://www.newsweek.com/2015/12/25/why-volkswagen-cheated-404891.html>.

[11] Goodman, McGrath, & Leah, (2016a, April 24). VOLKSWAGEN, EPA REACH 'CONCRETE PLAN' IN VEHICLE SOFTWARE CHEAT SCANDAL, *Newsweek.* Retrieved from <http://www.newsweek.com/buyback-offers-possible-fix-volkswagen-owners-deal-judge-450789>.

[12] Hakin, D. (2016, February 26), VW's Crisis Strategy: Forward, Reverse, U-Turn, *The New York Time.* Retrieved from <http://www.nytimes.com/2016/02/28/business/international/vws-crisis-strategy-forward-reverse-u-turn.html>.

[13]Kalaugher, L. (2015, October 29) Excess emissions from Volkswagen diesels could cause 59 early US deaths, *IOP Publishing.* Retrieved from <http://environmentalresearchweb.org/cws/article/news/63032>.

[14] Klier, T., & Linn, J. (2016). Comparing US and EU Approaches to Regulating Automotive Emissions and Fuel Economy.

[15]Kottasova, I. (2015, December 4). Volkswagen's sales are collapsing, *Cable News Network.* Retrieved from <http://money.cnn.com/2015/12/04/news/companies/volkswagen-sales-down/>.



- [16] LaReau, J., & Ryan, B. (2015, October 12). As crisis grows, Volkswagen dealers fear cash won't be enough, *Automotive News*. Retrieved from <http://www.autonews.com/article/20151012/OEM/310129960/as-crisis-grows-volkswagen-dealers-fear-cash-wont-be-enough>.
- [17] Le Page, M. (2015). How did Volkswagen rig emissions tests?. *New Scientist*, 227(3041), 11.
- [18] Mittelman, M. (2016, May 3). Sales Drop Is Smallest Since December, *BloobBerg L.P.* Retrieved from <http://www.bloomberg.com/news/articles/2016-05-03/volkswagen-brand-s-u-s-sales-drop-is-smallest-since-december>.
- [19] Moulson, G. (2016, April 13). Volkswagen says its managers' bonuses will be cut following the automaker's diesel emissions scandal, *U.S. News & World Report*. Retrieved from <http://www.usnews.com/news/business/articles/2016-04-13/volkswagen-to-cut-managers-bonuses-following-diesel-scandal>.
- [20] Rising, D., & Sopke, K. (2015, December 10). Volkswagen engineers worked on emissions cheating in U.S. as early as 2005, *MaineToday Media*. Retrieved from <http://www.pressherald.com/2015/12/10/volkswagen-engineers-began-working-on-cheat-in-2005-for-u-s-market/>.
- [21] Schlanger, Z. (2015, December 17). LETTER FROM ELON MUSK: VOLKSWAGEN'S CHEATING SCANDAL PUNISHMENT SHOULD BE TO MAKE ZERO-EMISSION CARS, *Newsweek*. Retrieved from <http://www.newsweek.com/letter-elon-musk-volkswagens-cheating-scandal-punishment-should-be-make-zero-406633>.
- [22] Selin, N., E. (2015b September 30). The company's deception could exceed \$100 million in costs from health damages, *Fortune*. Retrieved from <http://fortune.com/2015/09/30/volkswagen-diesel-scandal-consequences/>.
- [23] Selin, N., E. (2015a, September 30). Volkswagen emissions cheating caused \$100 million in health costs, according to analysis, *Grist*. Retrieved from <http://grist.org/climate-energy/volkswagen-emissions-cheating-caused-100-million-in-health-costs-according-to-analysis/>.
- [24] Sherk, J. (2014). Expand Employee Participation in the Workplace. The Heritage Foundation, March, 13.
- [25] Sky UK, 2016, *VW Sales Fall For First Time In 11 Years*, online retrieved on 8<sup>th</sup> May 2016 from <http://news.sky.com/story/1619342/vw-sales-fall-for-first-time-in-11-years>.
- [26] Sorokanich, B. (2016, April 21). VW Agrees to Buy Back U.S. Diesel Vehicles Affected by TDI Emissions Cheat, *Road and Track*. Retrieved from <http://www.roadandtrack.com/new-cars/car-technology/news/a28916/vw-tdi-emissions-settlement/>.
- [27] The Editorial Board, (2016, February 16). Beyond the Emissions Test at VW, *The New York Time*. Retrieved from <http://www.nytimes.com/2016/02/17/opinion/beyond-the-emissions-test-at-vw.html>.
- [28] Waecsh, M., H. (2016, January 8). Volkswagen Sales Drop for First Time in 13 Years, *Dow Jones*. Retrieved from <http://www.wsj.com/articles/volkswagen-group-sales-drop-for-first-time-in-13-years-1452275074>.
- [29] Yadigaroglu, I., (2015, December 18). A GREEN WAY TO PUNISH CHEATING VOLKSWAGEN, *Newsweek*. Retrieved from <http://www.newsweek.com/green-way-punish-cheating-volkswagen-406905>.

# Developing Web-Based Project Management Application for Event Organizer ABC

Bintoro Adi Guna  
Informatics Engineering  
Petra Christian University  
Surabaya, Indonesia

Alexander Setiawan  
Informatics Engineering  
Petra Christian University  
Surabaya, Indonesia

Rudy Adipranata  
Informatics Engineering  
Petra Christian University  
Surabaya, Indonesia

**Abstract:** ABC is an event organizer company based in Surabaya, Indonesia. Organizing events calls for project management. Unfortunately, ABC still does most of their business activities manually, which is prone to errors and mistakes such as miscalculations, loss of documents, and misinformation. To clear the vulnerability, ABC attempts to develop an application which fulfills their project management needs. The result of this research is a web-based application which is able to take and receive projects offers, assign staffs into groups, communicate with other staffs, create tasks, generate gantt tasks, create and realize project budgets, and sharing project documents.

**Keywords:** Event Organizer, Project, Project Management, CodeIgniter, Application

## 1. INTRODUCTION

ABC is a company based in Surabaya which has a division in entertainment and artist management industry. In addition to offering artists for entertainment purposes, ABC also takes offers from clients to organize events, which calls for project management. Unfortunately, most of ABC's project management activities are still done manually and are vulnerable to various mistakes and human errors such as loss of documents, miscalculations in project cost, and misinformation between project staffs.

Such vulnerabilities urged ABC to develop a project management application to fulfill their needs. Therefore, the author attempted to develop an integrated and user-friendly project management application containing essential features of project management required by ABC.

## 2. BASIC THEORY

### 2.1 Project Management

Kemp (2006) defines project as a dream with a deadline and a problem scheduled for solution [1]. Kemp (2006) also cited management in business as control [1]. If there is something uncontrolled, then there is no management in that business. From the preceding definitions, it can be concluded that project management is an act to ensure that the process of finding solution to a problem is kept under control. The term 'control' here means finished and punctual.

Project management components used in this application are as follows: scope management, time management, human resource management, cost management, and communications management.

### 2.2 MySQL

This application processes data inputted and outputted to and from MySQL databases. MySQL (My Structured Query Language) is a popular free and open-source database application. MySQL uses SQL scripting language which is a programming language adopted from English language and used to manipulate databases [3].

### 2.3 PHP

This application uses PHP for its server-side code in addition to HTML and CSS for the interface. PHP language is combined with CodeIgniter PHP framework to make development easier and faster.

PHP is an open-source programming language focused on server-side scripting, in which the codes are sent to server and executed there. Outputs from the codes are returned in the form of HTML to the client [2]. Because the codes are executed in the server, clients are not able to access the codes directly.

### 2.4 Model-View-Controller Architecture

As what has been discussed in point 2.3, CodeIgniter uses Model-View-Control (commonly referred to as MVC) architecture. The architecture splits applications into three objects: model for storing data in databases, view for the part of the application which is presented to the user, and controller which is the bridge between view and model. Controllers process input and output data given to or asked by users then place it in views for outputs or save it in models [4][5]. Figure 1 shows the MVC architecture.

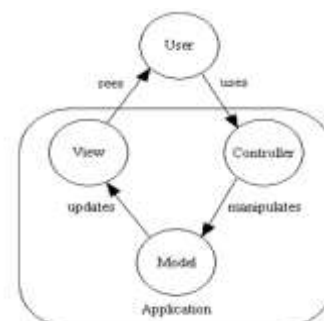


Figure 1 - MVC architecture

### 3. SYSTEM DESIGN

#### 3.1 Analysis of Existing Project Management System

Current project management system is done manually. This means that clients need to physically contact the company, either via phone, email, or coming directly to the office. After there's an agreement in scope and price, the company proceeds to assigning staffs to work on the project and creating a budget plan for the project. Company staffs assigned to work on the project would need to communicate by themselves, arrange their own schedules and tasks, and sharing documents on different locations. This keeps being done until the project is executed and completed.

#### 3.2 Problem Analysis of Existing Project Management System

Current project management system is cluttered and time-consuming especially for staffs when preparing for the project they are working on. They communicate by themselves, they schedule their own tasks, and they keep their documents online on different locations. This may lead to loss of documents, misinformation, and miscommunication between project staffs which may lead to the failure of a project. Project budgets being done manually may also lead to miscalculations and errors.

#### 3.3 Requirement Analysis of Existing Project Management System

In order to combat the risks and vulnerabilities pointed out in point 3.2, it can be identified that ABC needs an integrated project management application which is able to facilitate staff management, communications, and schedules. Additionally, ABC also requires the application to be able to facilitate budgeting process, from creating budgets to realizing them.

#### 3.4 Data Flow Diagram (DFD)

Data flow diagram (or DFD in short) is a graphic modelling technique showing the flow of data from external entities into the system, showed how the data moved from one process to another, as well as its logical storage [6].

##### 3.4.1 Context Diagram

Context diagrams in DFD shows the overall information and entity in an application. There are five entities involved in the application: artist, administrator, user, accountant, and staff.

##### 3.4.2 DFD Level 0

DFD Level 0 shows a more detailed process coming from the preceding context diagram. The name of the process involved in this application is Event Organizer Module.

##### 3.4.3 DFD Level 1

DFD Level 1 shows an even more detailed process coming from the preceding DFD Level 0. This level consists of three specific processes:

- Project registration, the act of registering projects based on requests from clients.
- Project budgeting and resource allocation, the act of allocating staffs (resources) and budgets. Additionally, this process also handles budget management.
- Interstaff communication, the act of allowing staffs to communicate and give tasks with each other.

Figure 2 shows the DFD Level 1 used to develop this application.

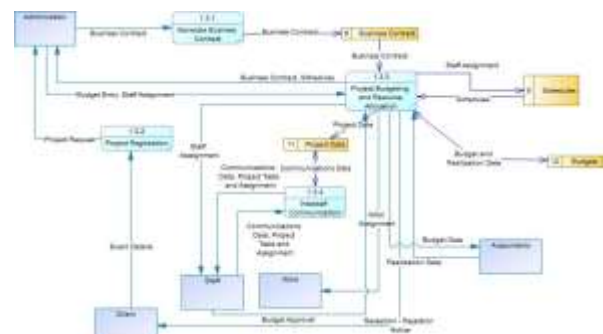


Figure 2 - DFD Level 1 of the application

#### 3.5 Entity Relationship Diagram (ERD)

Entity Relationship Diagram (or ERD in short) is a modelling technique which illustrates various entities in an information system as well as their relations to each other. The term 'entity' refers to physical items or existent concepts such as users, projects, et cetera. ERD can be considered as a foundation to build an application's database. The conceptual data model for the application can be seen in Figure 3.

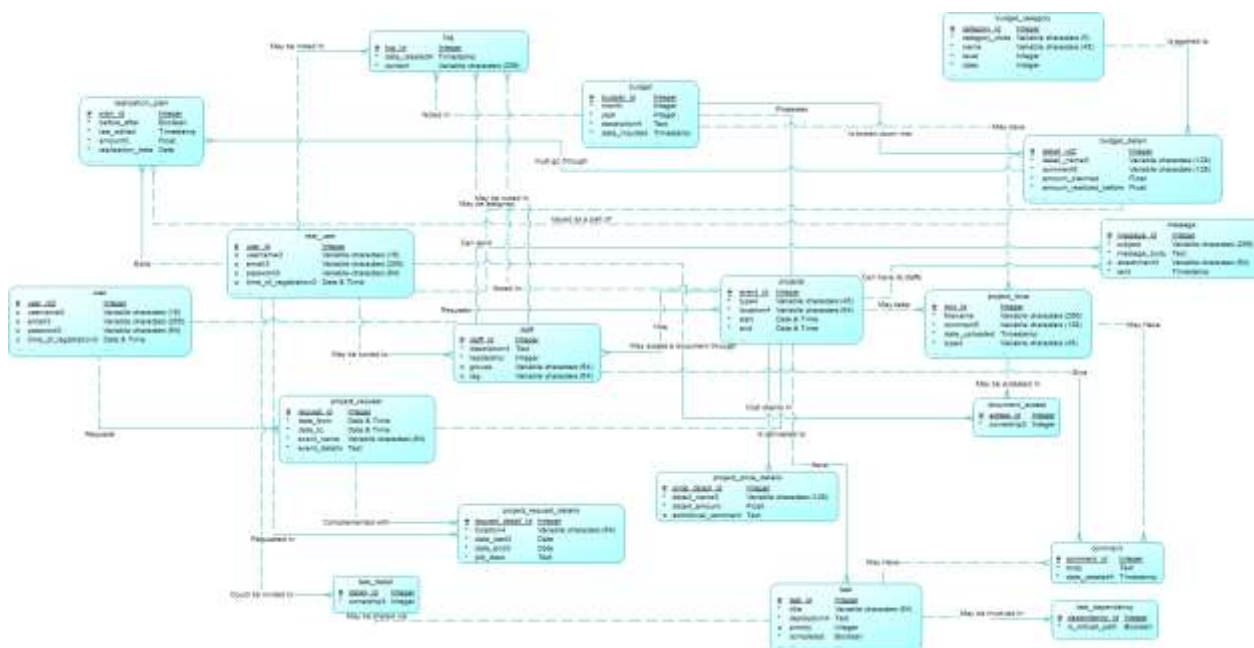


Figure 3 – Conceptual Data Model of the application

#### 4. APPLICATION IMPLEMENTATION

In this chapter the author is going to explain and discuss about the developed application. This chapter is divided into subchapter in which each is going to feature a short explanation and a screenshot of the feature.

##### 4.1.1 Sending Project Request

In order for a project to be recognized by the application, clients need to post their project requests by accessing the application’s homepage. Clients need to register for a free account before being able to post a request. Once registered and verified, clients need to head to a section called “Client Area” then press “Request A Project” button. Clients will then be taken through a form wizard to fill the details of their desired event. The form wizard is featured in Figure 4.



Figure 4 - Project request form wizard

##### 4.1.2 Responding to Project Requests



Figure 5 - Administration section log in page

Only users with admin privileges are allowed to see the project requests sent by client. Users will need to log in from the administration/back-end section of the application as seen in Figure 5. The privileged user will then be able to view and respond to project requests. Similar to clients requesting a new project, the user will also be guided via form wizard to respond the client’s request. The user will be allowed to directly reject the request, or put it on hold for later. The user will need to input essential data such as project price, performers, and documents (if available). The client will be notified via email about the development of his/her project request. Figure 6 features the form wizard that guides administrators in responding to a client’s project request.

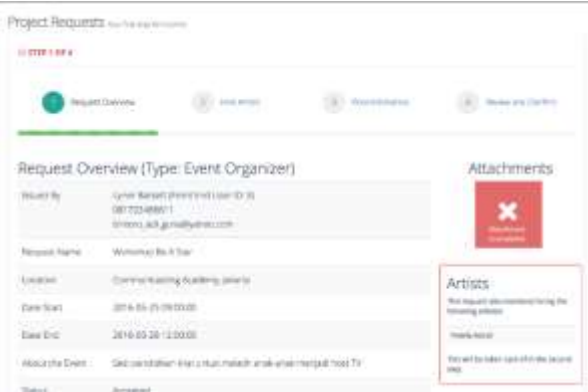


Figure 6 - Project response form wizard

Accepted project requests will be converted automatically to projects, allowing the administrators to assign staffs and create budgets for the corresponding project.

#### 4.1.3 Adding Resources and Budget

After approving a project request, administrators can immediately assign staffs and create budget for the newly-created project. Figure 7 shows the project information page which contains essential information such as project name, location, status, value, staffs, and link to budget. Administrators can also assign staffs from this page. Administrators can either register new staffs or select existing ones. Not only adding staffs, administrators will also need to configure a staff's leadership status; either the staff is a staff or is a project manager. Project managers have additional capabilities which will be discussed more in-depth in later subchapters.



Figure 7 - Project information page

Aside from adding staffs, administrators also need to create budgets for the project to make sure that the project cost does not exceed its value. Budget items are categorized into configurable categories. Administrators can determine whether each category belongs to income or expenditure. Figure 8 shows the interface of project budget creation page.

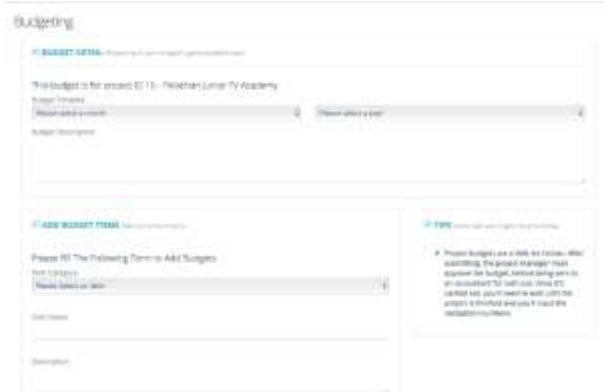


Figure 8 - Create new project budget

Once a project budget is created, it will need to be approved by the project's manager. Project managers are capable to either approve or disapprove the budget. If a manager chooses to do the latter, administrators will need to revise the budget. Project budgets must be approved. Otherwise, accountants will not be able to create realization plans, which, in reality, cash out the money required to pay for the project requirements.

#### 4.1.4 Managing The Budget

Once a budget is successfully approved by a project manager, accountants can immediately create realization plans to either cash in or out the money. Cash in process is simpler than cash out and can be conducted by both project managers and accountants. Cash in process is also faster. It is immediately realized once it is created. Cash out process, however, is a little bit complicated. Only accountants can create cash out realization plans and realize them. To create a plan, accountants need to choose which item they want to realize, then pick the realization date. Realization plans *cannot* be realized before the designated realization date. Once realized, however, the application will not only generate a cash in/out document, but will also record the realization in the project budget, which will be reflected in the "Budgeting" section in project managers' project management app, which will be discussed in the next subchapter. Figure 9 shows the interface of realization plan page used by accountants.

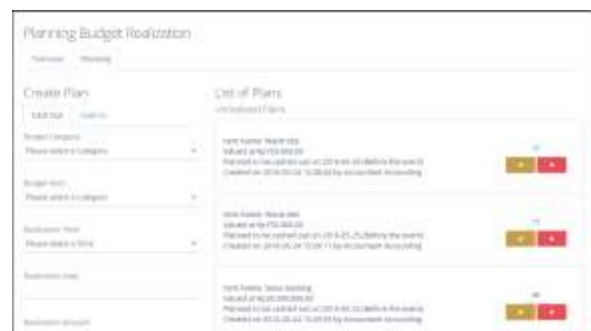


Figure 9 - Budget realization plan page

#### 4.1.5 Managing A Project

This subchapter discusses the core feature of the application, where staffs can collaborate to deliver high-quality, successful projects. The project management interface used by staffs have four panes while project managers have an additional pane: Budgeting. The following is the explanation of each pane.



#### 4.1.5.1 Project Overview



Figure 10 - Gantt chart generated in "Project Overview" pane

The Project Overview pane displays the essential information about the selected project such as project name, value, location, and execution date. In addition, it also displays the Gantt chart created by the project managers when available. The Gantt chart is powered by Google Charts and can be seen in Figure 10.

It is also possible to comment on the tasks within the Gantt chart. Users will need to switch from Gantt view to list view in order to view the task's details and post comments. This is due to the limitation of Google Charts. The author is not able to activate Google Chart's *onSelect* method in Google charts. Thus, the list view.

#### 4.1.5.2 Resources

The "Resources" pane involves staffs. Project managers can group their staffs from this pane in addition to tagging them. Created groups are immediately registered in the database. On the other hand, project managers can also remove staffs from groups and delete their tags. Staffs, however, do not share this privilege. Staffs can only see their fellow colleagues and their groups and tags.

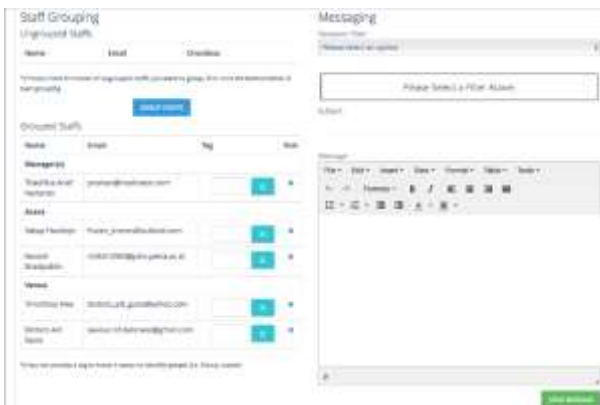


Figure 11 - Project resources pane

In addition to viewing, grouping, and tagging staffs, the Resources pane also allows staffs to send messages to other fellow staffs. Messages can also be broadcasted by choosing the appropriate filter then checking the appropriate checkboxes. There are four filters available:

- Tag, which means sending the message to only staffs with the selected tag(s).
- Group, which means sending the message to only selected group member(s).
- Individual, which means sending the message to only selected individual(s).
- Everyone, which means sending the message to every other staff.

Figure 12 shows the mailbox available to every project staff and a sample message.



Figure 12 - Staff inbox and sample message

#### 4.1.5.3 Budgeting

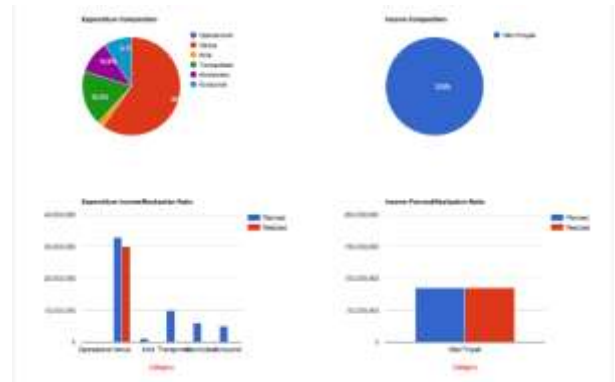


Figure 13 - Budgeting graphical information

The "Budgeting" pane allows project managers to view and control the usage of project budgets. Thus, this pane is only available to project managers. This pane consists of graphics informing the manager about the project's budget condition such as budget composition, planned/realization ratio as seen in Figure 13 in addition to budget information and list of budget documents. Project managers can also conduct a cash-in from this pane.

#### 4.1.5.4 Gantt/Tasks

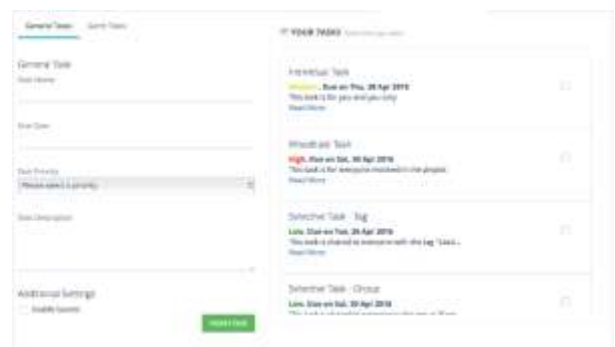


Figure 14 - Gantt/Task Pane

The "Gantt/Tasks" pane allows staffs to create tasks and project managers to create tasks which will be displayed in the Gantt chart. In addition to creating tasks for themselves, staffs can also share their tasks to other staffs. Shared tasks will be displayed on corresponding staffs' tasks pane. As for tasks displayed in Gantt chart, project managers can also task a certain group in the task. Please be noted, however, that a group must be created first before project managers can assign groups to the Gantt task. Figure 14 features the interface of Gantt/Task pane.

#### 4.1.5.5 Documents



Figure 15 - Documents pane

In “Documents” pane, as seen in Figure 15, staffs are allowed to upload and share documents related to the project. Shared documents will be displayed in the guests’ Documents pane and the guests will be able to view them. Documents shared in one project will not be displayed in other projects. In addition to uploading and sharing documents, staffs are also allowed to delete and view the documents. Viewing the document will open a new tab that will display the document and allow users to comment on it. However, the viewer will only be able to open PDF files. As for other file types, the user will be allowed to download the document. Document deletion is only available to document owners. Figure 16 shows the document viewer page.



Figure 16 - Document viewer page

## 5. CONCLUSION

Based on the results of application implementation, the author has concluded the following:

- The application is able to take project requests and respond to it, assigning staffs and creating budgets as well as managing them.
- The application is able to create budget realization plans and integrate their realization into the budget project to be viewed by project managers.
- Staffs are able to view their project schedules via Gantt chart, giving themselves and their colleagues tasks and comment on them, communicate via messages, and uploading documents as well as sharing, viewing, and commenting on them. All of those are done within the application.
- 80% survey respondents argued that the overall application is good while the remaining 20% respondents argued that the overall application is very good. From the survey results, it can be concluded that the application has met the requirements set by ABC company.

## 6. REFERENCES

- [1] Kemp, S. (2006). *Project Management Made Easy*. California: Entrepreneur Press.
- [2] *What is PHP?*. Retrieved May 2, 2015. <http://php.net/manual/en/intro-what-is.php>
- [3] Nixon, R. (2014). *Learning PHP, MySQL, Javascript, CSS & HTML5* (3rd ed). California: O’Reilly Media.
- [4] Qureshi, M. & Sabir, F. 2013. *A Comparison Model View Controller and Model View Presenter*. *Sci.Int.(Lahore)*, 25(1):7-9.
- [5] Kalelkar, M., Churi, P., Kalelkar, D. 2014. *Implementation of Model-View-Controller Architecture Pattern for Business Intelligence Architecture*. *International Journal of Computer Applications*, 102(12):16-21.
- [6] Data Flow Diagram (DFD)s: An Agile Introduction. Retrieved May 29, 2016. <http://www.agilemodeling.com/artifacts/dataFlowDiagram.htm>

# Making the Application of E-Marketplace with the Use of Virtual Account on Company

Hermawan Heriyanto  
Informatics Engineering  
Petra Christian University  
Surabaya, Indonesia

Alexander Setiawan  
Informatics Engineering  
Petra Christian University  
Surabaya, Indonesia

Adi Wibowo  
Informatics Engineering  
Petra Christian University  
Surabaya, Indonesia

---

**Abstract:** Indonesia is home to various marketplace sites such as Lazada, Bukalapak, and Tokopedia. The tight competition between marketplace sites encourages businesspeople to innovate in creating a new marketplace that is safe, comfortable and innovative. One innovative that can be done is of a system of cooperation with suppliers. Based on this background company trying to make an online application marketplace with a system of cooperation in a lease, or a combination of the rental fee and the fee. This application is expected to create security in the process of buying and selling as well as it can compete with the existing marketplace. The whole process of the transaction, the complaint and the ad would also be integrated into the system, thereby reducing the risk of human error. Overall the application complies with the requirements of the company. In addition to getting the ease in getting customers, suppliers also feel the ease and safety in working in company

**Keywords:** Marketplace, Rent, Fee, System Cooperation

---

## 1. INTRODUCTION

Indonesia is home to various marketplace sites such as Lazada, Bukalapak, and Tokopedia. These Marketplace offers an advantages to maintain the position in order to remain competitive with competitors. For example one of the benefits of cooperation in Tokopedia is registration for free. These advantages encourages businesspeople to innovate in creating a new marketplace to compete with the marketplace that has been stood.

One of the innovative in establishing new marketplace is to create a different system of cooperation with suppliers. The system of cooperation may take the form of rent, fees and a combination of rental and fees.

The company that wants to create a whole new marketplace with a category of goods is a bag. In this case, company plan to create an online marketplace of applications are supported by the implementation of the joint account system and transaction fee. This application is expected to reduce the sense of insecurity customers and capable of competing with the marketplace that have stood because can facilitate collaboration with various suppliers of bags.

## 2. BASIC THEORY

### 2.1 E-Marketplace

E-Marketplace is a virtual market where buyers and sellers meet and conduct various types of transactions. Here the process of exchange of goods, services, money or information. Virtual marketplace connecting sellers and buyers via the Internet. The process of exchange of services and goods in the virtual world is defined as e-commerce. There are 5 things that infrastructure to support e-commerce, namely people, public policy, marketing and advertising, support services, and business partnerships. [4]

E-commerce also has several characteristics, namely: the Internet is a media intermediary in the process of buying and selling, the process of the transaction and the exchange of goods, information and services. Besides e-commerce also has several types, namely: Business to Business (B2B), Business to Consumer (B2C), Consumer to Consumer (C2C), Consumer to Business (C2B) and Collaborative Commerce.

Marketplace, including the type of Business to Business (B2B). This is because the transactions that occur in the marketplace involves the relationship between the seller and not directly in the hands of the end (end-user) / customer. [2]

## 3. SYSTEM DESIGN

### 3.1 Requirement Analysis

Reasons doneonline marketplace creation of applications in this research is to create an innovative online marketplace with the three types of application of the system of cooperation with PHP and CSS Design [1][3]. Rental system facilitates cost provider of goods where payment is always the same regardless of the number of sales that occurred. Fee system of cooperation in which the provider of goods will pay some perasetase goods sold and a combination of rental and fee system where the system will help suppliers in making the transition from the system to the rental fee.

### 3.2 System Analysis

In this research, the application program that is used to apply 3 system of cooperation with suppliers. The application of the registration process starts from supplier to sales calculation process that has been accomplished.

#### 3.2.1 Supplier Registration

On the supplier registration process, the application data to be recorded by the system supplier and activation email sent to the supplier. Supplier account activation process begins with checking the data stored on the system and activate the account if achieved compliance. Explanation of supplier registration process can be seen in Figure 1



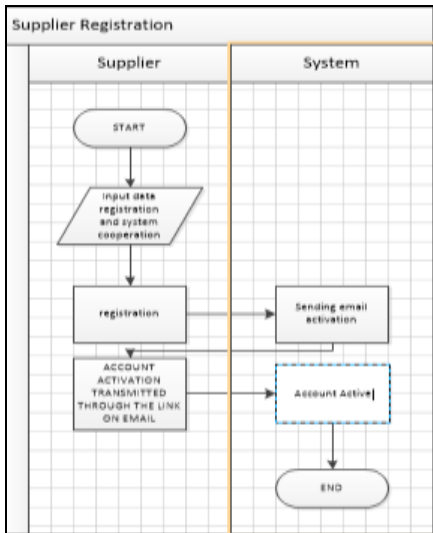


Figure 1 Supplier Registration

### 3.2.2 Sales of Goods

#### 3.2.2.1 Ordering Goods

In the process of ordering goods, the data of goods ordered and the shipping costs will be recorded and the recording system will provide confirmation of an order made by the customer. Explanation of the process of ordering goods can be seen in Figure 2.

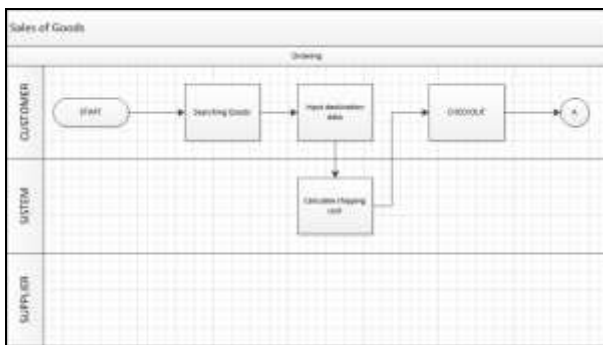


Figure 2 Ordering Items

#### 3.2.2.2 Customer Payment

In the process of payment, payment data will be recorded and admin company will provide confirmation of payment to the customer. After admin confirmation, the system will provide notification to the supplier that the booking has been paid by the customer. Explanation customer payment process can be seen in Figure 3

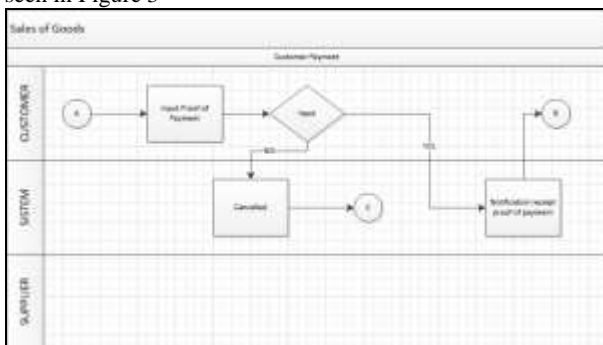


Figure 3 Customer Payment

#### 3.2.2.3 Shipping

In the shipping process, shipping of data will be stored and notification every stage of the delivery of goods will be accepted by the customer. Explanation of the shipping process can be seen in Figure 4.

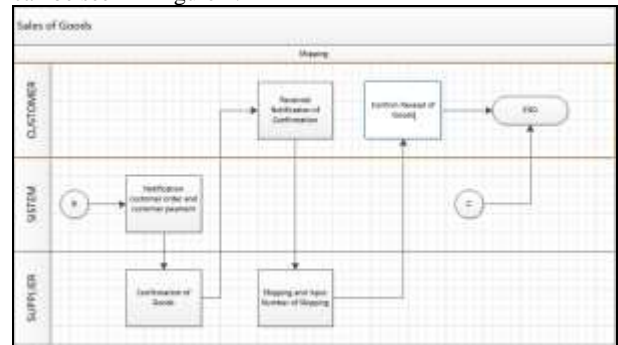


Figure 4 Shipping

#### 3.2.3 Calculation of sales that have been completed

In this process, a notification will be sent details of the sale have been calculated for a range of dates specified. Furthermore, the supplier must confirm that the details have been received and true as in Figure 5.

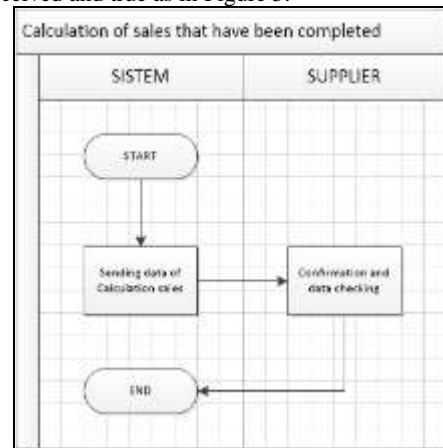
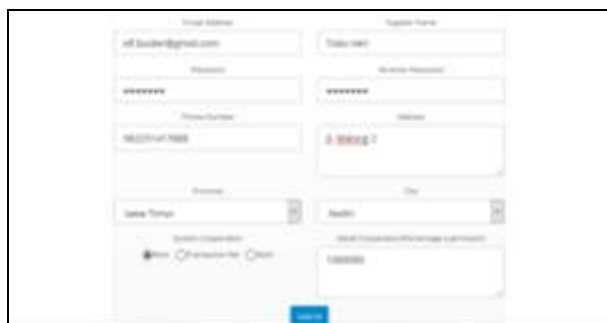


Figure 5 Calculation of Sales

## 4. APPLICATION IMPLEMENTATION

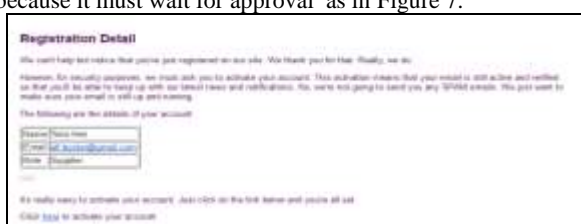
### 4.1 Implementation of Supplier Registration

In this application, the supplier can register your account by completing the registration form account as shown in Figure 6. For example, email is filled with elf.buster@gmail.com, the password is also required, the name Heri stocked shop, filled phone number 082231417888, address filled jl.malang 2, select the province of East Java, select the town of Kediri, a system of cooperation been rent and filing requests cooperation filled Rp 1,000,000.



**Figure 6 Supplier Registration**

After filling and signing up is complete, the supplier will receive an email containing the account activation link to activate your account supplier but still can not be used because it must wait for approval as in Figure 7.



**Figure 7 Email Activation**

Once active, the company will contact the applicant and make determinations cooperation. Admin of the company can classify the supplier in accordance with the agreement as shown in Figure 8, and activate the account so that the supplier can use an account that has been registered in accordance with Figure 9.



**Figure 8 Classify Supplier Category**

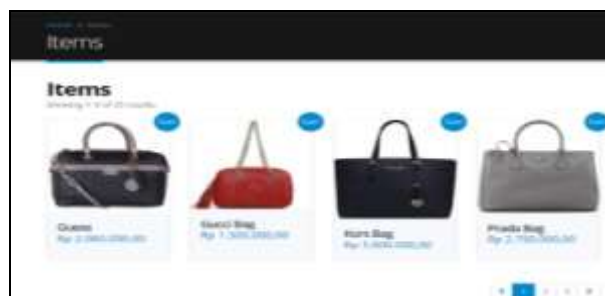


**Figure 9 Activation Supplier Account**

## 4.2 Sales of Goods Implementation

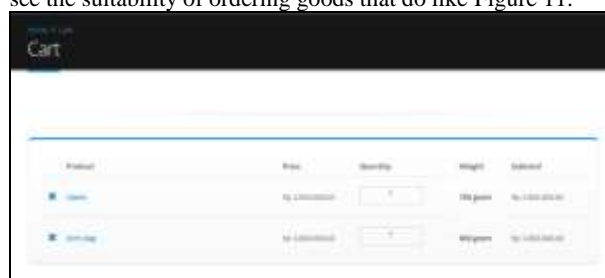
### 4.2.1 Ordering

Customers who visit the marketplace application is able to see a wide variety of goods sold as Figure 10 without the need to perform the login process



**Figure 10 List of Items**

However, if you want to order goods by including a shopping cart, login process needs to be done first. After entering all the items to be purchased on the shopping cart, the customer can see the suitability of ordering goods that do like Figure 11.



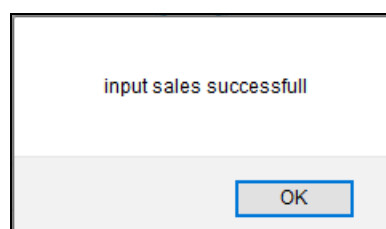
**Figure 11 Shopping Cart**

The data required to be filled out completely and clearly so that the process of calculating the cost of shipping and delivery process runs smoothly. For example, for a field filled with Hermawan name, phone number field filled 08223141788, the address field is filled with jl.malang 2, East Java provinces were selected field, field Kediri city selected. Upon completion of the complete data, by pressing the button updates the shipping cost shipping fees will be calculated as shown in Figure 12.



**Figure 12 Destination and Shipping Data**

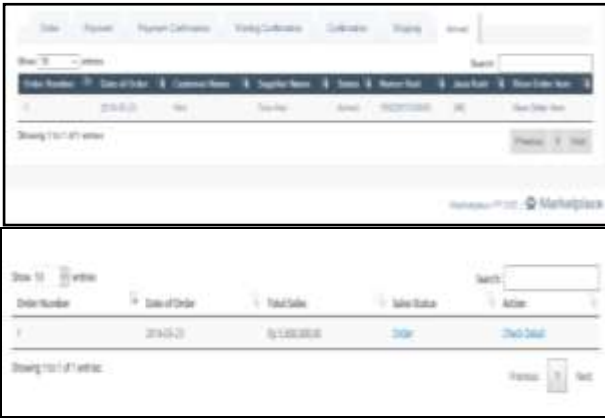
After making the checkout process, the customer will get a notification purchases of goods as successful as in Figure 13.



**Figure 13 Notification Successful**

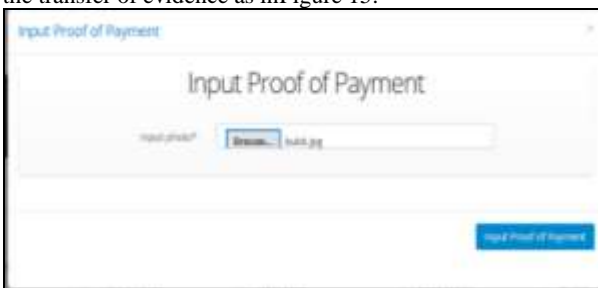
### 4.2.2 Customer Payment

Customers who have made the process of buying goods, can see the status of orders as in Figure 14.



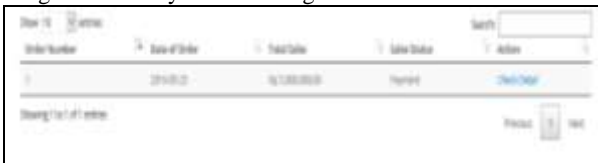
**Figure 14 List of Sales**

For customers who have not made the payment process, can make the payment process by the method of transfer and input the transfer of evidence as in Figure 15.



**Figure 15 Input Proof of Payment**

After making a payment, purchase status will change to payment as Figure 16 and then customer only waiting to see the goods delivery status like Figure 17.

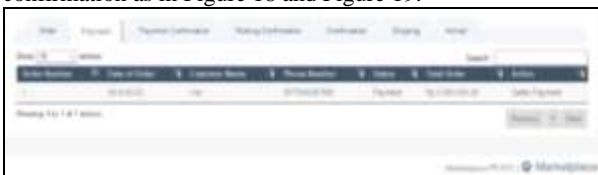


**Figure 16 List of Payment Sales**

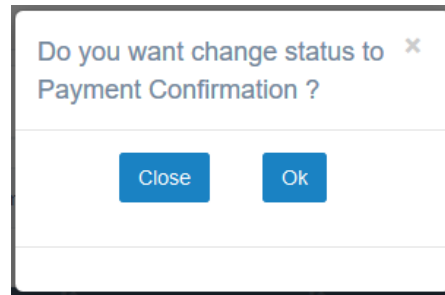


**Figure 17 Status Delivery for Customer**

Confirmation process customer payment can be made by admin company if the customer has done inputting proof of transfer by changing the status of payment into a payment confirmation as in Figure 18 and Figure 19.



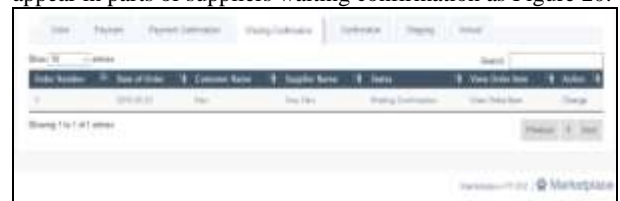
**Figure 18 List of Payment for The Company**



**Figure 19 Changing to Payment Confirmation**

### 4.2.3 Shipping

Sales have been confirmed payment by admin company will appear in parts of suppliers waiting confirmation as Figure 20.



**Figure 20 List of Waiting Confirmation Sales**

Suppliers are required to confirm the booking made by the customer and confirmed the results will be displayed as in Figure 21.



**Figure 21 List of Confirmation Sales**

Information will also be given confirmation to the customer such as Figure 22.



**Figure 22 Information about Confirmation for Customer**

The next item is ready to be delivered by courier JNE and the supplier will perform a data inputting shipping information such as Figure 23.

Figure 23 Input Data Shipping

Information delivery with shipping details will also be given to the customer as Figure 24.

Figure 24 Information Shipping for Customer

If the goods have arrived at the destination, the supplier will change the status of the sale be arrived like Figure 25 and Figure 26

Figure 25 Changing Status to Arrive

Figure 26 List of Sales Which Status is arrived

### 4.3 Calculation of Sales

The process of inputting payment begins by selecting a date range of sales calculation has been completed and has not been paid to the supplier. For example, from 1 May 2016 to 25 May 2016. Figure 27 is the result of the calculation details of sales during the specified time period.

Figure 27 Calculation of Sales

Details of the sale can be sent like Figure 28. Details will be sent as an email to the supplier as in Figure 29.

Figure 28 Sending Calculation of Sales

Figure 29 Email of Detail Calculation

## 5. CONCLUSION

Based on the test results we can conclude the following:

- Applications are made in accordance with the plan of the system to be applied to the company
- System of cooperation with suppliers is divided into three:
  - Rent: the fee paid is always fixed every month
  - The combination of rental and fees: fees paid by a percentage of the sales that have occurred coupled with the cost of rent per month
- Process of Sales divided into 3 phases, among others order, payment, and shipping
- The whole process of the transaction, the complaint and the advertisement will be integrated in the system, thereby reducing the risk of human error

## 6. REFERENCES

- [1] Duckett, J. (2011). HTML & CSS Design and Build Websites. Indiana: John Wiley & Sons, Inc
- [2] Laudon, K.C., dan Traver, C.G. (2013). E-commerce (9<sup>th</sup> ed). Pearson.
- [3] Nixon, R. (2014). Learning PHP, MySQL, Javascript, CSS & HTML5 (3rd ed). California: O'Reilly Media.
- [4] Turban, E. (2012). Electronic Commerce (7<sup>th</sup> ed). Pearson.