Research in Collective Intelligence through Horse Racing in Hong Kong

LYU1604
Lam Cheuk Hei (1155049403)
Chan Lok Wai (1155050024)
Supervised By Prof. Michael R. Lyu

Abstract

Horse racing is a popular sport in Hong Kong, many people love betting on it. In the situation of Hong Kong, although many people always lose money when they gambling on horse racing, there are still many people so talented on predicting horse racing, they won a lot of money, they even won in consecutive races, which is really hard to do that by random prediction, we believe that they have the ability to choose the right horses, and win in the future. In this project, we want to develop a system that can collect the choices of those people who always win, and then give a suggestion to our clients, of course, the suggestion is come from the collective intelligence. As soon as our clients created the account of our website, they can predict horses through our website, they can see our suggestions, but they are not necessary to follow our suggestions if they think they have a better choice, and in the back end of our server, we collect clients' predictions, and keep updating the suggestions based on our clients' predictions.

In this project, it is an experiment trying to find out the power of collective intelligence, we want to know how accurate it is, and we want to know if it can help people to make money.

Contents

1 Introduction

1.1 Motivation

Crowdsourcing is a hot topic in recent years. A crowdsourcing system helps people to solve problem by the data provided by user. However, will users input some fake data that hinder the task? Is there any way to filter the inappropriate data? Collective intelligence is one of the application in crowdsourcing. It involves consensus, social capital and formalisms such as voting systems, social media and other means of quantifying mass activity. Under quantifying mass activity, we can measure the data provided by user easily. Since our group desire to know the effectiveness of crowdsourcing, we are going to conduct an experiment, to test whether the wisdom of crowd is better than decisions made by single member. We choose horse racing as our research base because it provides discrete result, unlike stock market, and it is wide variety. These factors help us to judge the credibility of each users.

1.2 Objective

We are going to design a voting system to let user predict horse racing result. After collecting the data from user, we will analyse the data by our algorithm, considering the win rate and balance of each user. Finally, we will provide some tips, a list of hot horse, for user. We hope to design a good algorithm to prove the value of collective intelligence. So that we can release the application and attract people to use in the next semester.

1.3 Background

The concept of collective intelligence originated in 1785 with Marquis de Condorcet, whose "jury theorem" states that if each member of a voting group is more likely than not to make a correct decision, the probability that the highest vote of the group is the correct decision increases with the number of members of the group. Another precursor of the concept is found in entomologist William Morton Wheeler in 1911. He observed the collaborative process at work in ants that acted like the cells of a single beast, seemingly independent individuals can cooperate so closely as to become indistinguishable from a single organism.

Collective intelligence appears in sociobiology, political science. Although this term is not originated from computer science, computer is the most suitable tool because it involves a lot of data. There are also some examples in the internet using this concept.

Wikipedia is one famous example, it is a free online encyclopedia that allows users to edit any article, all contents are collected from users. Besides, Facebook is also a commonly used example, it allows people to have and share an experience with other users. However, these example is difficult to verity the power of collective intelligence because users can share anything freely without analysis.

To verify the wisdom of crowd, we choose horse racing to be our research base. In this semester, we have designed a web application with the functions of log-in, creating account, voting, ranking, and etc. This will be one of the tools for us to collect data and to find out the power of collective intelligence.

1.4 Definition

Definition of collective intelligence

Two or more people independently acquire information and this information is combined and processed through social interaction, which provides a solution to a specific problem such that it cannot be implemented by isolated individuals (adapted from Krause et al 2009).

Definition of crowdsourcing

Although it is hard to define crowdsourcing exactly, we can define it as a general-purpose problem-solving method: a crowdsourcing system invites a crowd of people to help solving a problem (Doan et al. 2011). Estellés-Arolas and Gonzáles-Ladrón-de-Guevara (2012) gave us another detailed definition:

Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken.

Both of these definitions is usable for the purposes of our project.

2 Technical Support

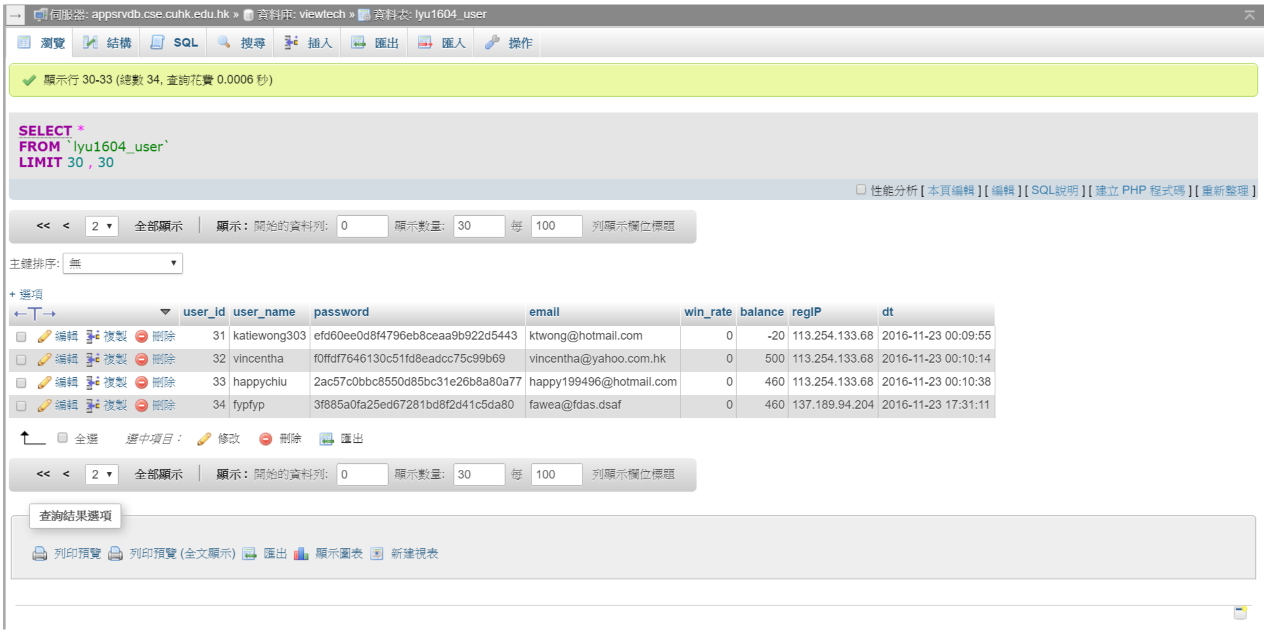
2.1 Development Environment

Our system is a web application which involves back-end and front-end development. For the back-end development, we used the database by phpMyAdmin to manage the data. For the front-end development, we used Dreamweaver to design our webpage and some text editors (e.g. Sublime Text) for coding. During development, we have tested our application on different web browsers, including Google Chrome, Firefox, Safari, Internet Explorer, Microsoft Edge. It ensures our application can work on most of the environment.

2.2 Software tool: Database by phpMyAdmin

Since our application involves data collection, data analysis and log-in system, to store the data and user information, a database is needed. We have chosen a MySQL database by phpMyAdmin provided by the VIEW Lab. It is a user-friendly platform for developing and managing our database.

We have created some tables on phpMyAdmin to store different data, including 'user' (Figure 1.) to store the user informaion, 'race' to store the racing information, 'buy_record' to store the horse name predicted by users, and etc. We will discuss the architecture of our database on detail in section 4.1.



伺服器: appsrvdb.cse.cuhk.edu.hk » 資料庫: viewtech » 資料表: lyu1604_user
瀏覽
結構
SQL
搜尋
插入
匯出
匯入
操作

✓ 顯示行 30-33 (總數 34, 查詢花費 0.0006 秒)

```
SELECT *
FROM `lyu1604_user`
LIMIT 30 , 30
```

□ 性能分析 [本頁編輯] [編輯] [SQL說明] [建立 PHP 程式碼] [重新整理]
<< < 2 ▾ 全部顯示 | 顯示： 開始的資料列: 0 顯示數量: 30 每 100 列顯示欄位標題
主鍵排序: 無 ▾
+ 選項

←T→	user_id	user_name	password	email	win_rate	balance	regIP	dt
□ 編輯 複製 刪除	31	katiewong303	efd60ee0d8f4796eb8ceaa9b922d5443	ktwong@hotmail.com	0	-20	113.254.133.68	2016-11-23 00:09:55
□ 編輯 複製 刪除	32	vincentha	f0ffdf7646130c51fd8eadcc75c99b69	vincentha@yahoo.com.hk	0	500	113.254.133.68	2016-11-23 00:10:14
□ 編輯 複製 刪除	33	happychiu	2ac57c0bbc8550d85bc31e26b8a80a77	happy199496@hotmail.com	0	460	113.254.133.68	2016-11-23 00:10:38
□ 編輯 複製 刪除	34	fypfyp	3f885a0fa25ed67281bd8f2d41c5da80	fawea@fdas.dsaf	0	460	137.189.94.204	2016-11-23 17:31:11

□ 全選 選中項目： 修改 刪除 匯出
<< < 2 ▾ 全部顯示 | 顯示： 開始的資料列: 0 顯示數量: 30 每 100 列顯示欄位標題
查詢結果選項
列印預覽
列印預覽 (全文顯示)
匯出
顯示圖表
新建視表

Figure 2.1 A part of user table

After the tables are created, we need to input data to some tables, for example the race information and race result collected from the Jockey Club. It is impossible to input data one by one since there are lots of records. MyphpAdmin provides a function, ‘import’, to let us input multiple records in .csv format. We can copy all racing information on next raceday, save it in one table and import all data once to the database. It is much more convenient for the back-end development.

2.3 Software tool: Dreamweaver

For the front-end development, we used Dreamweaver as our software tool. It is a proprietary web development tool developed by Adobe Systems. We used it to design our webpage instead of other software because it combines a visual design surface known as Live View and a code editor with standard features. Using the preview function, we can see the effect immediately after editing some code. The Live View is also complete with live Javascript, database queries, server-side code. We can test the webpage with our back-end system simultaneously.

Besides, the function ‘Check Page’ is very useful for developing the webpage. We need to ensure every functions can work, and every design look good on every browser. The ‘Check Page’ function allow us to check for browsers compatibility regularly. It is more convenient for testing than opening the webpage on different browsers.

Moreover, since our webpages are in PHP format, we cannot direct open it on browsers. Upload to the department server is one of the way, but it is trouble to upload it many times after every editing. Dreamweaver allows us to run and preview PHP codes within the software. It helps us to save a lot of time for uploading the webpage.

3. System Architecture

3.1 Overall Architecture

The overall system contains a server part and client part. In the server part, it connects to the database, which can store all information of our system. Our server becomes a medium of intercourse between the users and database, when a user input something on the website, the server can get the information and send it to the database.

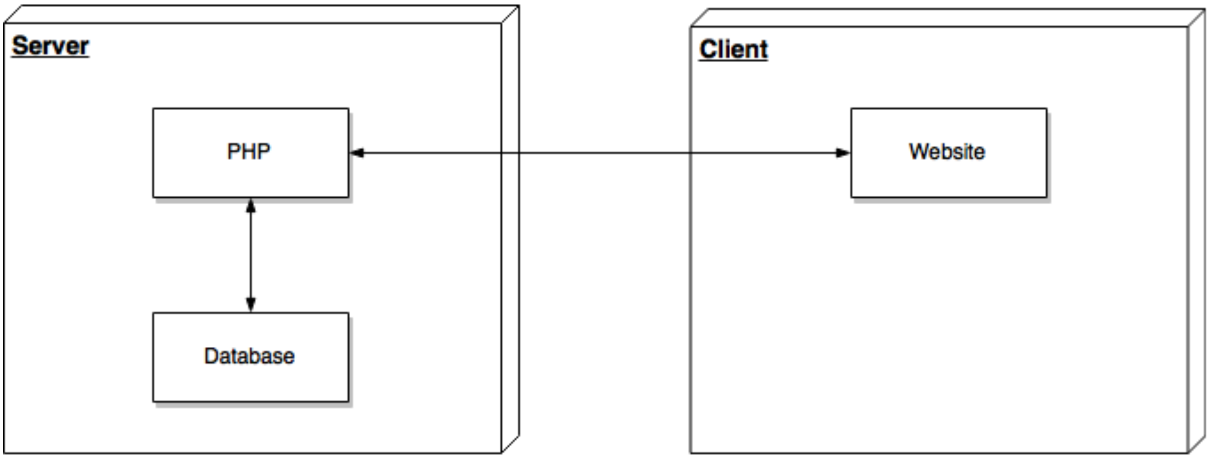


Figure 3.1 A graph of the overall architecture of our system

3.2 System Overview

In this part, we will focus on our system in more detail by showing the full process of predicting horse.

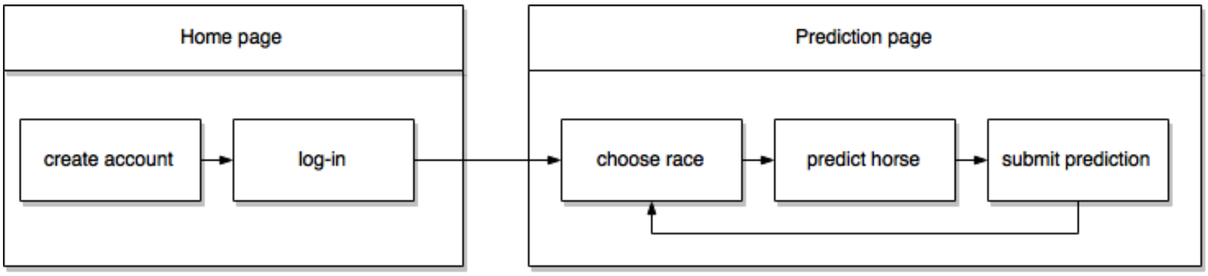


Figure 3.2 The flow of predicting horse racing

First of all, our users need to create an account on the website, they need to provide the user name, email address and the log-in password. If there are other users who used the same user name or email.

After the user successfully created the account, the user can log-in to our web page, because only the users who logged in can use the prediction function.

At this moment, our users are ready for the prediction, users can go to the prediction page, and then choose the race, in most of the time, there are more than 8 races in one racing day, users can choose it depending their knowledge to different races.

Users can see the information of that race, such as horses and the corresponding jockey and trainer, also the current odds for every horses. We provide users to predict two types in a race, one is “win”, and the other is “place”. When the user picks “win” for a horse, that means the user thinks that horse will win the race. When the user picks “place”, that means the user thinks that horse will be the first three in the race. It is obvious that users can only choose one “win” and they can choose a maximum of three “place”. After submitting their prediction, our server would record it. Then users can predict other races, and repeat the above process. That’s the whole process of predicting horse racing.

Beside the prediction function, we designed some features for users to let them have a better experience of using our website. When the user successfully created an account, we give them $500 virtual money, users can use those moneys to predict horse racing, they may win money and lose money, they can check their current balance. Also, they can check their win rate, it is calculated by the server depending on their correct predictions and wrong predictions, we will discuss it in more detail in the later part of this report.

3.3 Activity Diagram

In this part, we will talk about the activity of each process in our system.

Create Account

To create account, users need to submit a user name, password and email address, and our server would check our database, to see if there is other user using the same user name or email. If someone is already using that user name or email address, users need to input the above information again. If no one uses that user name or email, this account can be created.

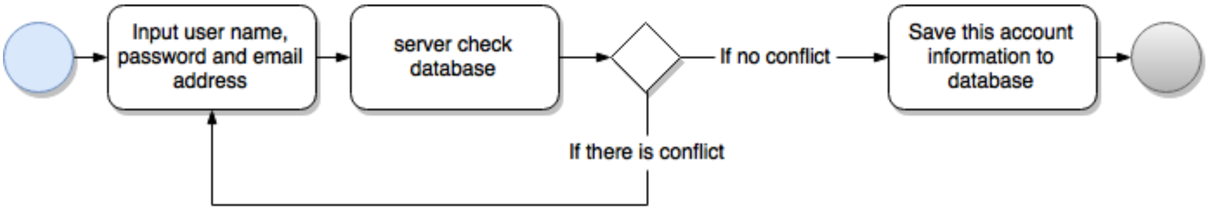

Input user name, password and email address
server check database
If no conflict
Save this account information to database
If there is conflict

Figure 3.3 Activity diagram showing the flow of creating account

Log-in

When users click the log-in button on our page, the client side would start the log-in function. The web page would go to the log-in page, and there are boxes for users to type the user name and password. The web then gets these inputs and send them to server. The sever then calculate the md5 of the password, and send a query to database to check if the combination of user name and md5 password is in the database. Then the server will send message to the web page, it would reload again to show successful log-in or show an error message if something is wrong.

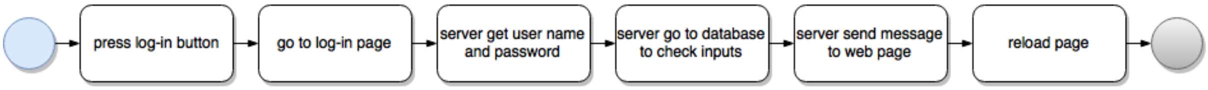

press log-in button
go to log-in page
server get user name and password
server go to database to check inputs
server send message to web page
reload page

Figure 3.4 Activity diagram showing the flow of log-in

Log-out

When users press the log-out button, the web page would delete the user cookies, including the user name, user id, and then reload the page for guess user.

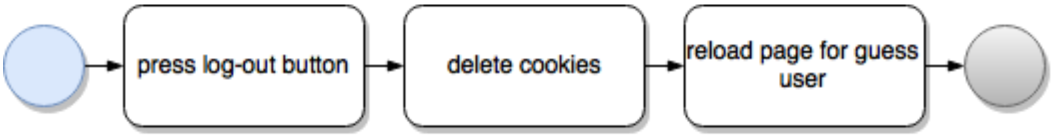


Figure 3.5 Activity diagram showing the flow of log-out

Check Log-in Status

Every time when users go to another page, we do a checking of log-in status, we check out the cookies, if cookies are not empty, server would go to database to ensure the user name and password are correct. If the user has already logged-in, the page would keep logging-in. Otherwise the page is in guess user mode.

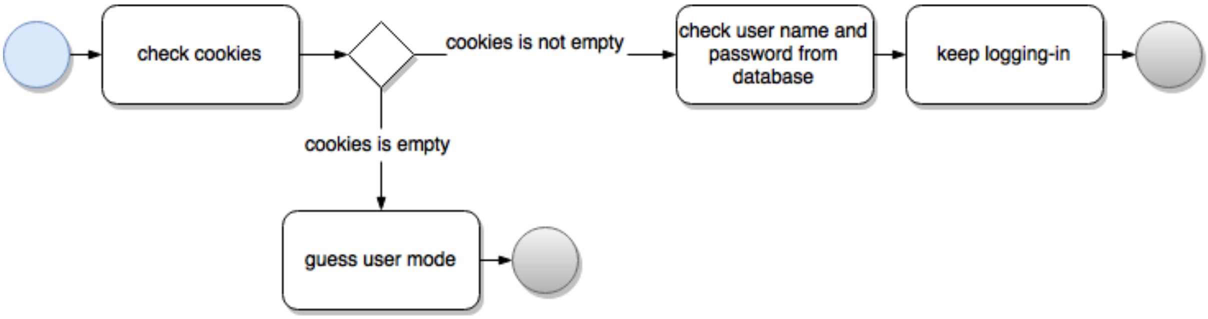


Figure 3.6 Activity diagram showing the flow of checking log-in status

Predict Horse Racing

To predict horse racing, as mentioned, users need to log-in, and then go to prediction page to choose a race, and select their prediction. Also, this prediction would save to our database.

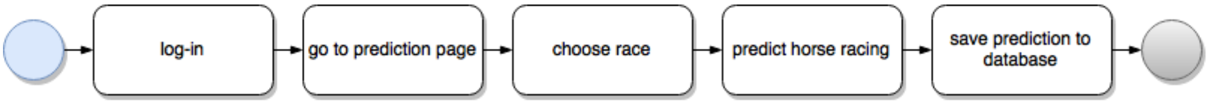


Figure 3.7 Activity diagram showing the flow of predicting horse racing

Check Balance

First of all, users need to log-in to our website, and then go to user information page, and then the page would show the balance from our database.

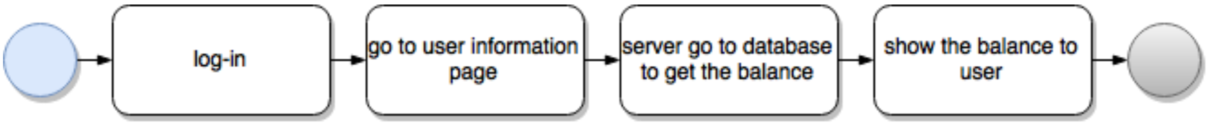

log-in
go to user information page
server go to database to get the balance
show the balance to user

Figure 3.8 Activity diagram showing the flow of checking balance

4 Design and Implementation

4.1 Server Side

4.1.1 Database: Schema

This is the schema of our database:

user(user_id: integer, user_name: varchar, password: text, email: varchar, win_rate: double, balance: double, regIP: varchar, date_time: timestamp)

buy_record(buy_record_id: int, buy_date_time: timestamp, user_id: int, amount: int, race_date: date, race_id: int, win: text, place1: text, place2: text, place3: text, earn_money: double)

User Table

Each tuple is a unique account. It stores the user name, password in md5, the email address, win_rate of the user, current balance, the IP address of the user, and the date and time of the creation.

Attribute	Description	Data type
user_id	The unique id of account	Integer(Auto increment)
user_name	The name of account	Varchar
password	The md5 password of the user's password	Text
email	Email address of the user	Varchar
win_rate	The win_rate of the user	Double
balance	The amount of money that the user has	Double
regIP	IP address of the user's device	varchar
date_time	Record the time of account creation	timestamp

Buy record Table

Every tuple is a unique record. Buy_record_id is the primary key of this table. User_id is a foreign key.

Attribute	Description	Data type
buy_record_id	The unique id of record	Integer(Auto increment)
buy_date_time	The date and time of that record created	timestamp
user_id	The user_id of the buyer	int
amount	It is the amount of predicting horse racing, each choice costs $10, and maximum is $40(1win and 3 place).	int
race_date	The date of horse race	date
race_id	The race number of that racing day	int
win	The name of horse which predicted to win	text
place1	The name of horse which predicted to be first three	text
place2	The name of horse which predicted to be first three	text
place3	The name of horse which predicted to be first three	text
earn_money	The money that user earns by accurately predicting horse raing, zero if all predictions are wrong	double

4.1.2 Implementation:

PHP and SQL queries

PHP is very popular is nowadays, it is a general-purpose scripting language which is suitable to the server side on web server. It can connect to our database, which is very useful for us. In the following we will talk about different parts of the PHP.

Setup connection to database:

```
// MySQL connection
$conn=new mysqli(DB_SERVER, DB_USER, DB_PASSWORD, DB_DATABASE);
// Check connection
if ($conn->connect_error) {
   die("Connection failed: " . $conn->connect_error);
}
```

Insert informaion to database - Create account:

```
<?php

$user_name = $_POST["user_name"];
$password = $_POST["password"];
$email = $_POST["email"];

// MySQL connection
$conn=new mysqli(DB_SERVER, DB_USER, DB_PASSWORD, DB_DATABASE);
// Check connection
if ($conn->connect_error) {
   die("Connection failed: " . $conn->connect_error);
}

$sql = "INSERT INTO user (user_name, password, email, balance) VALUES ('" .
$user_name ."','" . $password ."','" . $email ."','500')";
```

```
if ($conn->query($sql) === TRUE) {
    echo "New account created successfully";
} else {
    echo "Please use another user name or email";
}
$conn->close();
?>
```

Insert informaion to database - Create buy record:

```
<?php

$buy_date = $_GET["buy_date"];
$buy_time = $_GET["buy_time"];
$user_id = $_GET["user_id"];
$amount = $_GET["amount"];
$race_date = $_GET["race_date"];
$race_id = $_GET["race_id"];
$win = $_GET["win"];
$place1 = $_GET["place1"];
$place2 = $_GET["place2"];
$place3 = $_GET["place3"];

// MySQL connection
$conn=new mysqli(DB_SERVER, DB_USER, DB_PASSWORD, DB_DATABASE);
// Check connection
if ($conn->connect_error) {
    die("Connection failed: " . $conn->connect_error);
}

$sql = "INSERT INTO buy_record (buy_date, buy_time, user_id, amount, race_date,
race_id, win, place1, place2, place3) VALUES ('" . $buy_date ."','" . $buy_time ."','" .
$user_id ."','" . $amount ."','" . $race_date ."','" . $race_id ."','" . $win ."','" . $place1 ."','" .
$place2 ."','" . $place3 ."')";

if ($conn->query($sql) === TRUE) {
   echo "New record created successfully";
} else {
   echo "Error: " . $sql . "<br>" . $conn->error;
}

$conn->close();
?>
```

For another functions which are similar to the above PHP code, the followings are the SQL query:

Functions	SQL query
update user balance	UPDATE user SET balance='$balance' WHERE user_name='$user_name'
update user win rate	UPDATE user SET win_rate='$win_rate' WHERE user_name='$user_name'

We have shown how to insert informaion to database, and we will talk about how to get data from our database.
Get user information:

```
<?php
   // MySQL connection
   mysql_connect(DB_SERVER, DB_USER, DB_PASSWORD);
   // Database connection
   mysql_select_db(DB_DATABASE);
   // Rereieve user informaion
   $data=mysql_query("select * from user where user_name={$_SESSION['usr']}");
```

```
  $row2 = mysql_fetch_assoc(mysql_query("SELECT user_id,user_name, email,
win_rate, balance FROM user WHERE user_name='{$_SESSION['usr']}'"));
  ?>
```

For another functions which are similar to the above PHP code, the followings are the SQL query:

Functions	SQL query
get the user name, balance and win_rate whose blanace are top 10 of all users	SELECT user_name, balance, win_rate FROM `lyu1604_user` ORDER BY balance DESC limit 10
get the race information	select * from lyu1604_race_table where race_no = '$race_no'

4.1.3 Data security

SQL Injection Attack

It is a very serious problem, users attack the database from the input. If the server do not check the input, the information in database are not protected, they can steal the information or even delete the information. For example, the user may input 1’ or ‘1’ = ‘1 in the user name, and just type something in the password box, and then they can log-in the syatem if we don not check this input. The reason why this will happen is that this input causes the SQL query changed, here is the example:

Original:

```
SELECT user_id,user_name FROM lyu1604_user WHERE user_name='1’ or ‘1’ = ‘1'
AND password='".md5($_POST['password'])."'
```

Becomes:

```
SELECT user_id,user_name FROM lyu1604_user
```

This is how SQL injection works, so that we need to do something in the server to check the user input, we do not allow them to log-in if the user name contains some special characters. The following code shows that we checked the input before we get data from database.

```
    $_POST['username'] = mysql_real_escape_string($_POST['username']);
    $_POST['password'] = mysql_real_escape_string($_POST['password']);
    $_POST['rememberMe'] = (int)$_POST['rememberMe'];

    // Escaping all input data

    $row = mysql_fetch_assoc(mysql_query("SELECT user_id,user_name FROM
lyu1604_user WHERE user_name='{$_POST['username']}' AND
password='".md5($_POST['password'])."'"));

    if($row['user_name'])
    {
      // If everything is OK login

      $_SESSION['usr']=$row['user_name'];
      $_SESSION['id'] = $row['user_id'];
      $_SESSION['rememberMe'] = $_POST['rememberMe'];

      // Store some data in the session

      setcookie('tzRemember',$_POST['rememberMe']);
    }
    else $err[]='Wrong username and/or password!';
```

PHP function to prevent SQL injection attack

4.2 Front-end

4.2.1 Webpage

The title of our web application is ‘King of Tipsters’. Our webpages are written in .php format. PHP is a general-purpose scripting language that is especially suited to server-side web development. Since our webpage generally connects to the database, we choose to use this scripting language to develop our webpages. In addition, it can be used in combination with HTML code, so we can combine our design and function of the system.

Before connecting to the database, we use HTML5, CSS and JavaScript to design the whole interface. HTML5 and CSS are used to design the layout of the webpages. The HTML <div> element is always used together with CSS in our website to design the typesetting. For the visual style of the webpage, CSS is mainly used, for example the background color, font style, border of box, and etc. Besides, JavaScript is worked on some functionalities, for example the back to top button, the menu bar.

We have tested that our webpage supports most of the web browsers which are commonly used, for example Google Chrome, Windows Edge, Internet Explorer 11, Firefox, Safari.

4.2.2 User Interface Design

Our web application is mainly seperated in three webpages, the home page, the login page, the prediction page. They are all linked together by the menu bar. Here is the sitemap of our webpages:

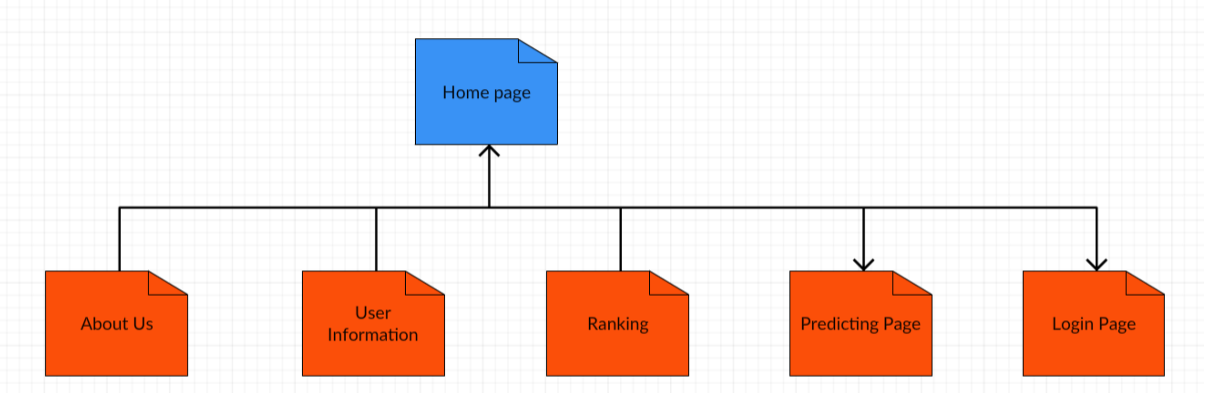


Figure 4.1 Sitemap of the webpages

For our home page, we apply an one page parallax HTML design which is clean and functional. User can get all the information they want in the homepage, for example race information, user information, win rate, balance and ranking. Figures 4.1 - 4.4 are some screen captures from the home page, 4.5 is the login and register page and 4.6-4.7 are the prediction page.

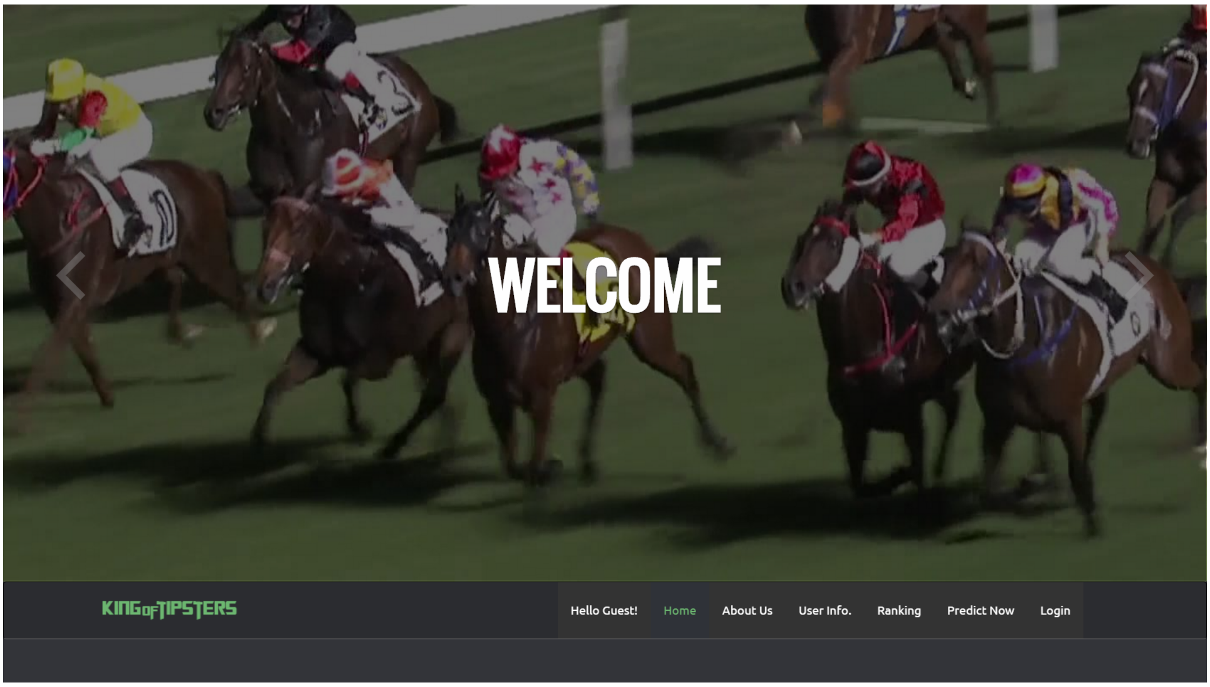
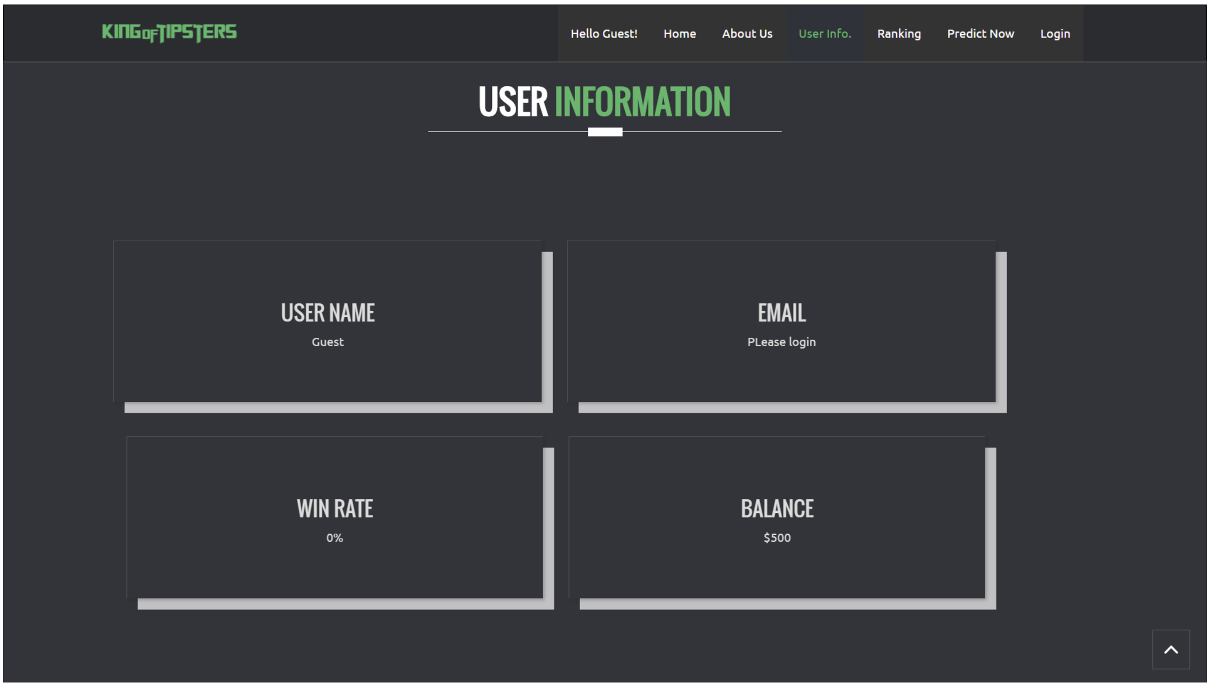


Figure 4.2 Home Page with the Menu Bar



USER INFORMATION

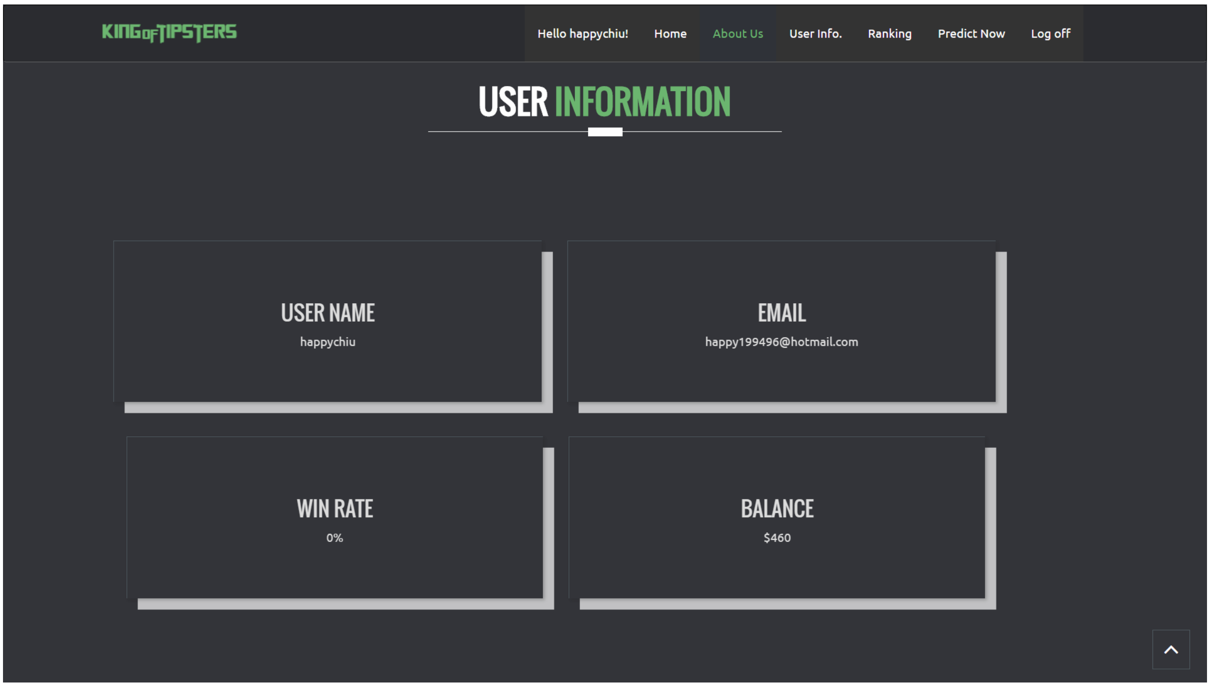
USER NAME
Guest

EMAIL
PLease login

WIN RATE
0%

BALANCE
$500

Figure 4.3 User Information berfore Login



USER INFORMATION

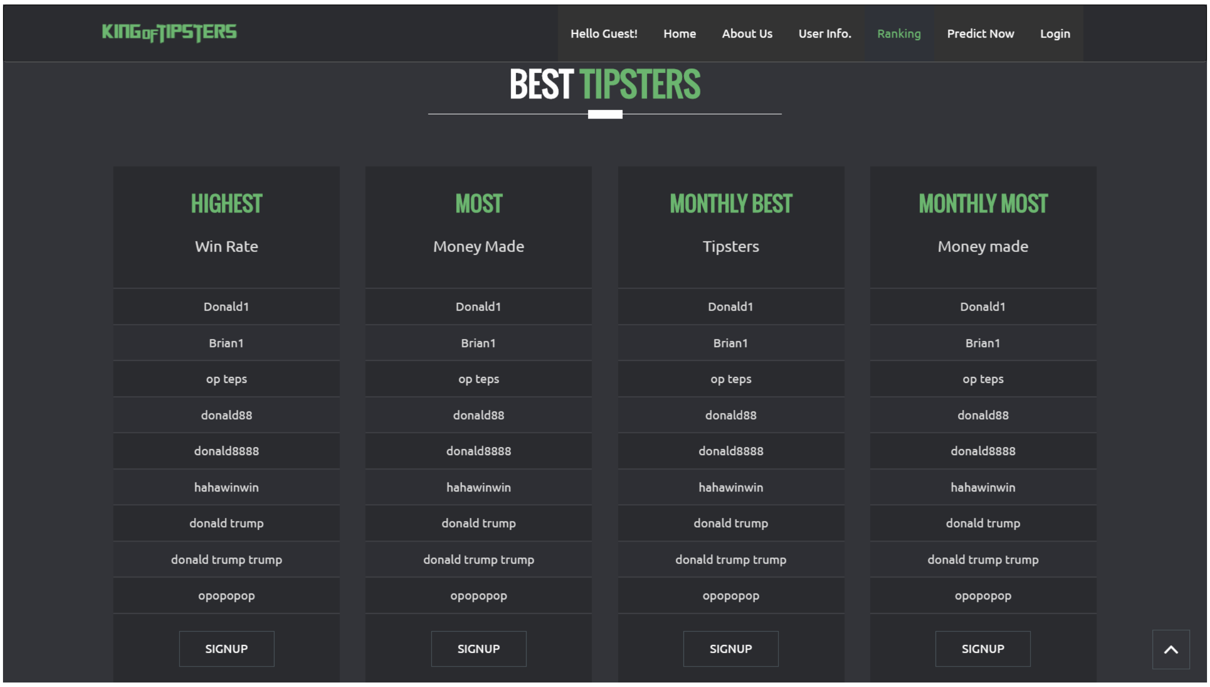
USER NAME
happychiu

EMAIL
happy199496@hotmail.com

WIN RATE
0%

BALANCE
$460

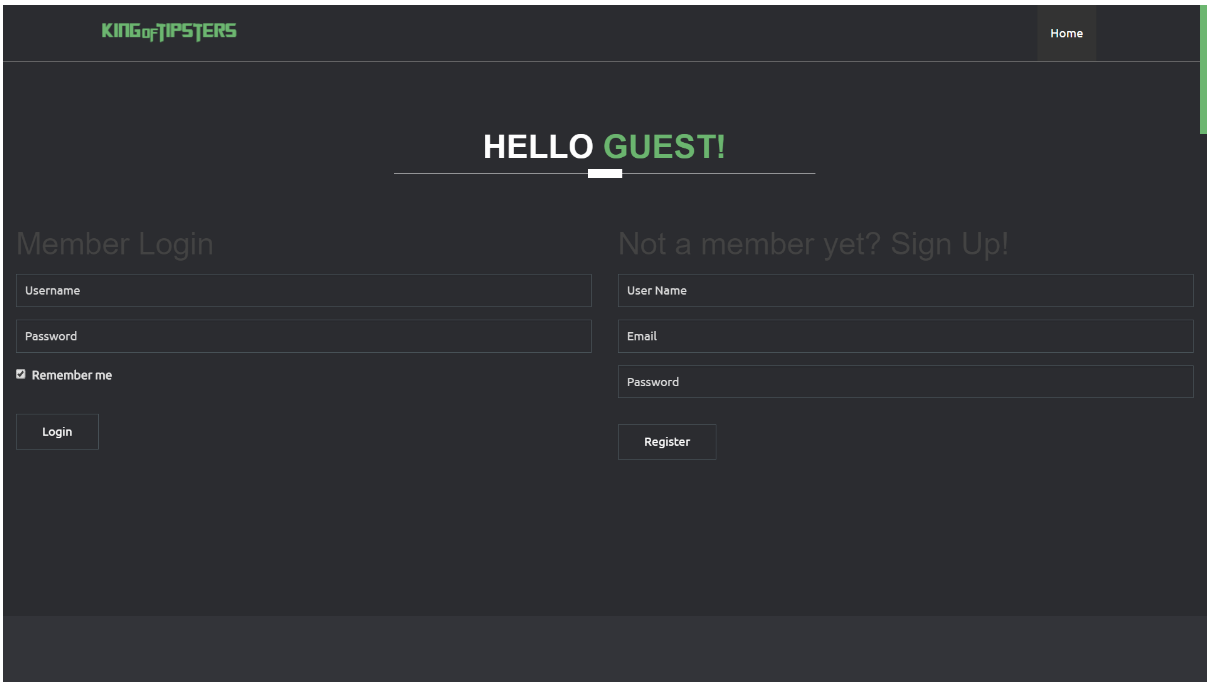
Figure 4.4 User Information after Login



BEST TIPSTERS

HIGHEST Win Rate	MOST Money Made	MONTHLY BEST Tipsters	MONTHLY MOST Money made
Donald1	Donald1	Donald1	Donald1
Brian1	Brian1	Brian1	Brian1
op teps	op teps	op teps	op teps
donald88	donald88	donald88	donald88
donald8888	donald8888	donald8888	donald8888
hahawinwin	hahawinwin	hahawinwin	hahawinwin
donald trump	donald trump	donald trump	donald trump
donald trump trump	donald trump trump	donald trump trump	donald trump trump
opopopop	opopopop	opopopop	opopopop

Figure 4.5 Ranking



HELLO GUEST!

Member Login

Username

Password

☑ Remember me

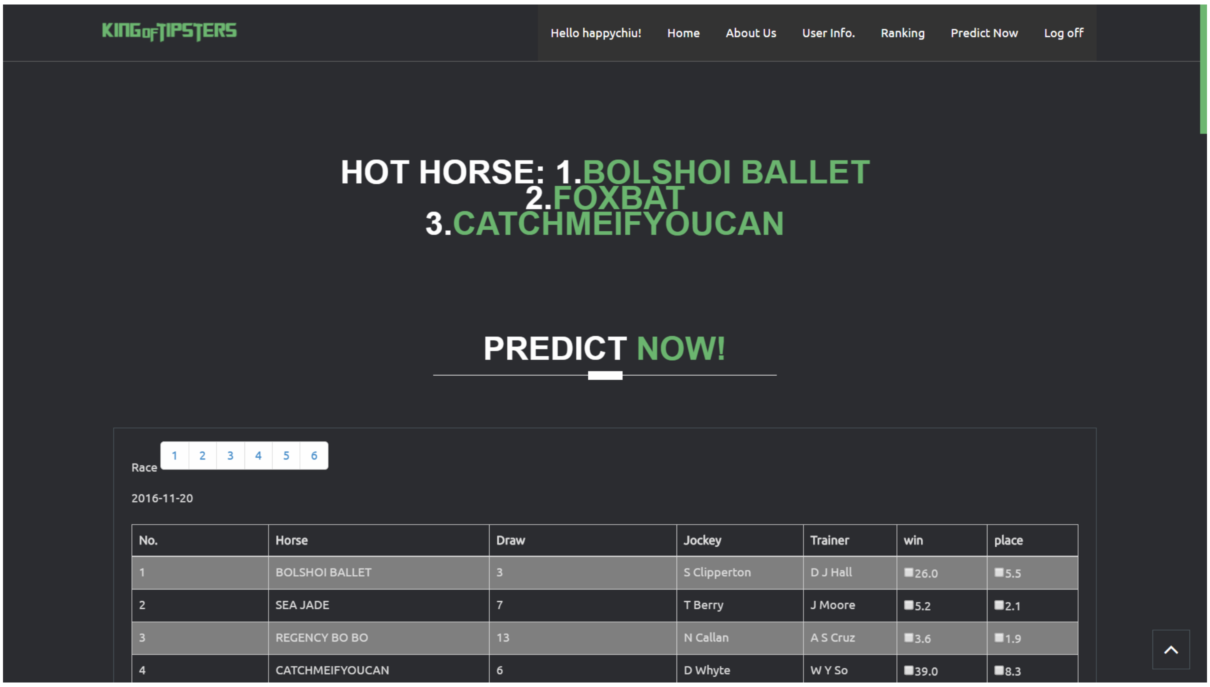
Not a member yet? Sign Up!

User Name

Email

Password

Figure 4.6 Login and Register Page



HOT HORSE: 1.BOLSHOI BALLET
2.FOXBAT
3.CATCHMEIFYOUCAN

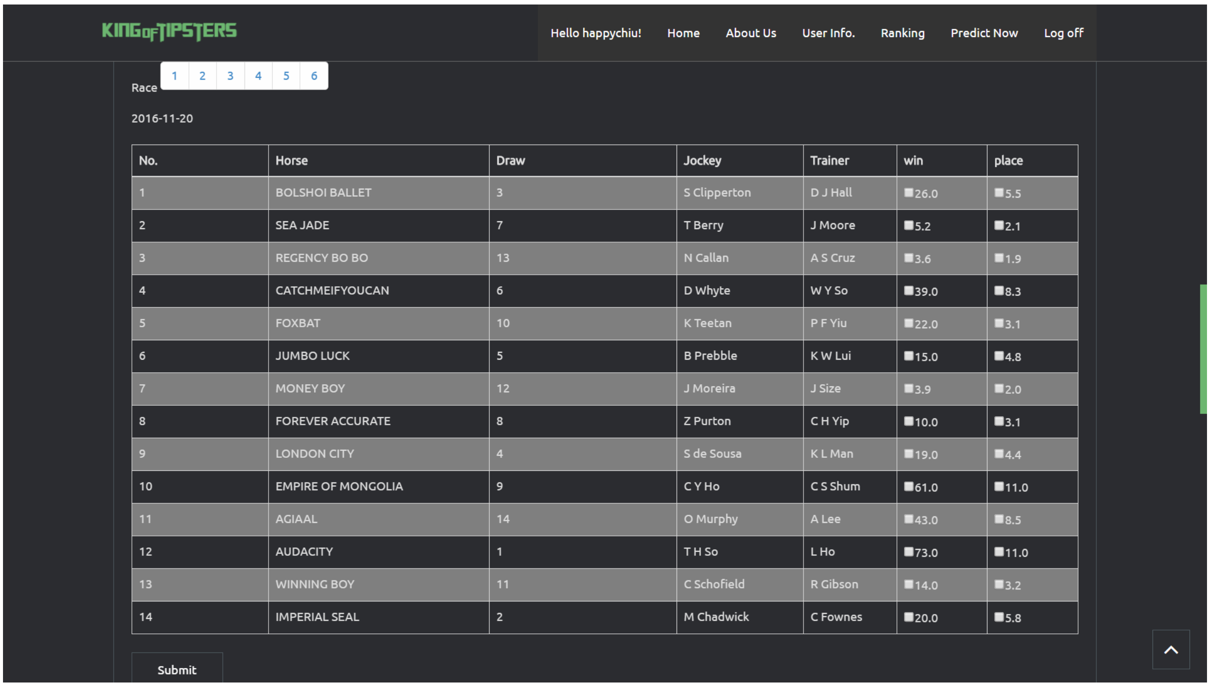
PREDICT NOW!

Race 1 2 3 4 5 6

2016-11-20

No.	Horse	Draw	Jockey	Trainer	win	place
1	BOLSHOI BALLET	3	S Clipperton	D J Hall	☐26.0	☐5.5
2	SEA JADE	7	T Berry	J Moore	☐5.2	☐2.1
3	REGENCY BO BO	13	N Callan	A S Cruz	☐3.6	☐1.9
4	CATCHMEIFYOUCAN	6	D Whyte	W Y So	☐39.0	☐8.3

Figure 4.7 Prediction Page



2016-11-20

No.	Horse	Draw	Jockey	Trainer	win	place
1	BOLSHOI BALLET	3	S Clipperton	D J Hall	□26.0	□5.5
2	SEA JADE	7	T Berry	J Moore	□5.2	□2.1
3	REGENCY BO BO	13	N Callan	A S Cruz	□3.6	□1.9
4	CATCHMEIFYOUCAN	6	D Whyte	W Y So	□39.0	□8.3
5	FOXBAT	10	K Teetan	P F Yiu	□22.0	□3.1
6	JUMBO LUCK	5	B Prebble	K W Lui	□15.0	□4.8
7	MONEY BOY	12	J Moreira	J Size	□3.9	□2.0
8	FOREVER ACCURATE	8	Z Purton	C H Yip	□10.0	□3.1
9	LONDON CITY	4	S de Sousa	K L Man	□19.0	□4.4
10	EMPIRE OF MONGOLIA	9	C Y Ho	C S Shum	□61.0	□11.0
11	AGIAAL	14	O Murphy	A Lee	□43.0	□8.5
12	AUDACITY	1	T H So	L Ho	□73.0	□11.0
13	WINNING BOY	11	C Schofield	R Gibson	□14.0	□3.2
14	IMPERIAL SEAL	2	M Chadwick	C Fownes	□20.0	□5.8

Figure 4.8 Predict Table in Prediction Page

4.2.3 Functionality

4.2.3.1 Registration

Before predicting and getting horse tips, users need to create an account of the website. This action can be done in the login and register page (Figure 4.5.). To create an account, user can follow this way:

1. Click on the Login button on the menu bar

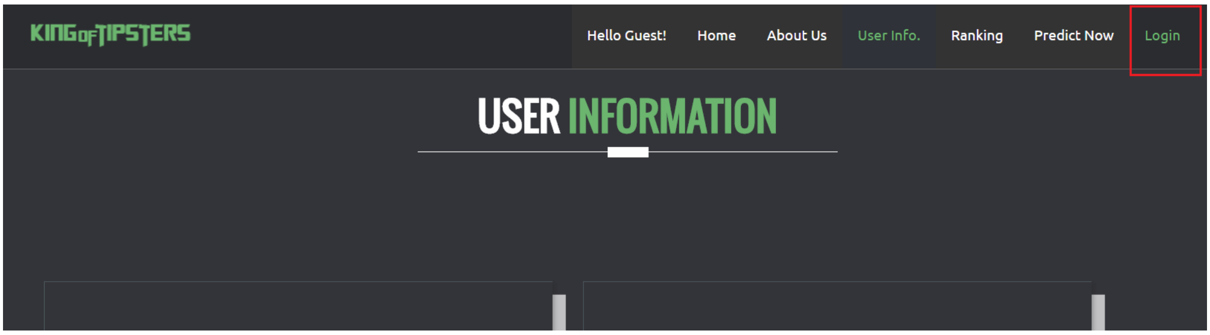


Figure 4.9 Login Button

2. Input the user information, including User Name, Email and Password
3. Click on the Register button

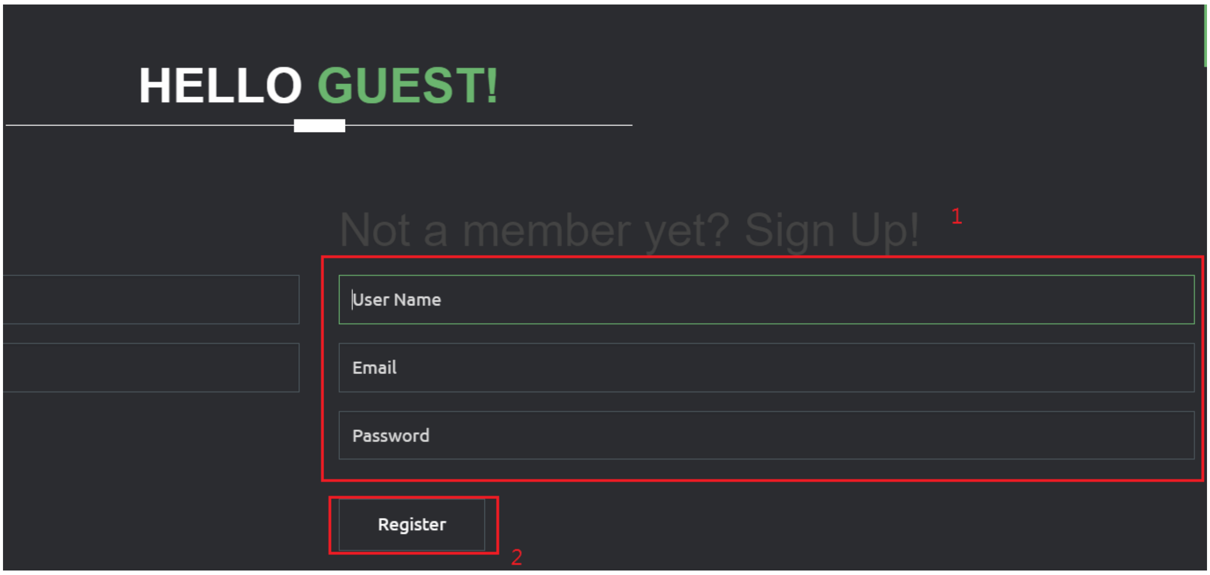
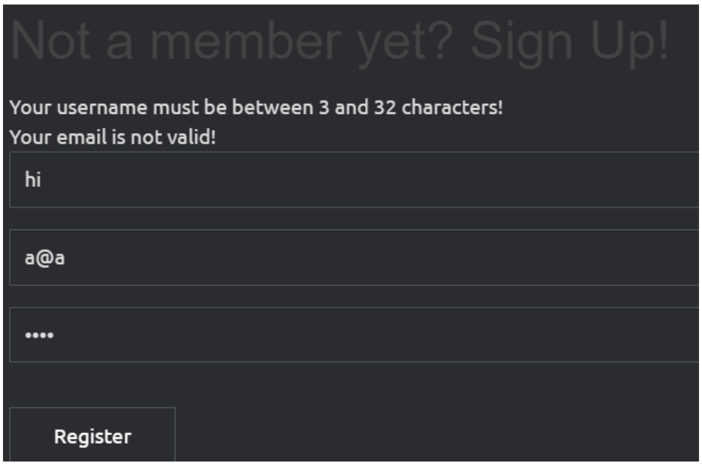


Figure 4.10 Register Area

4. If there is no problem for the information inputted, the account will be created susscessfully. However, if the username is not between 3 and 32 characters, or it is already taken by other user, the account will not be created and the error syntax will be shown on the register page. Besides, we also do not accept invalid email format.



Not a member yet? Sign Up!

Your username must be between 3 and 32 characters!

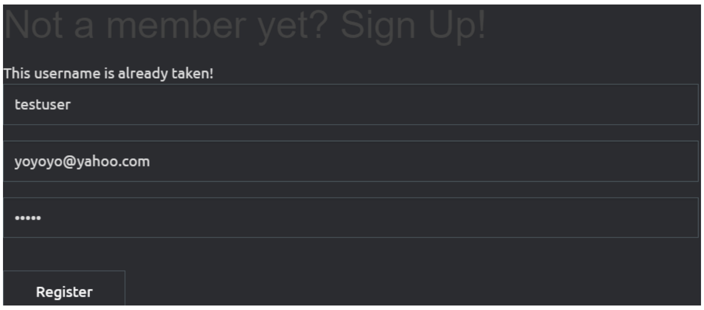
Your email is not valid!

hi

a@a

••••

Figure 4.11 Create account failed Demonstration1



Not a member yet? Sign Up!

This username is already taken!

testuser

yoyoyo@yahoo.com

•••••

Figure 4.12 Create account failed Demonstration2

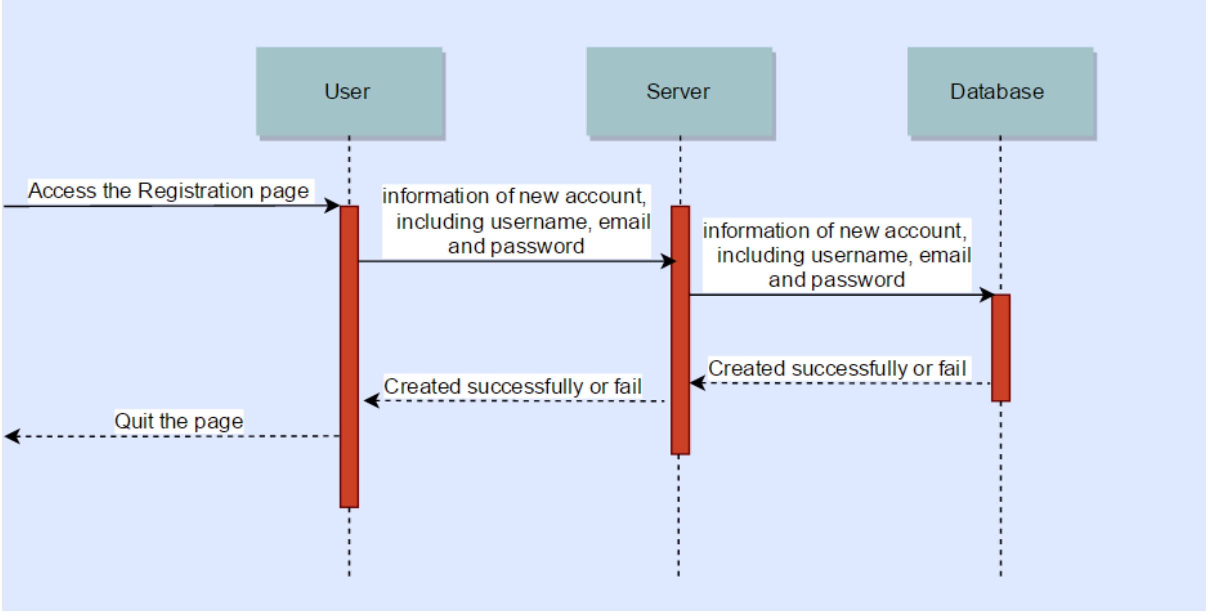

User
Server
Database
Access the Registration page
information of new account, including username, email and password
information of new account, including username, email and password
Created successfully or fail
Created successfully or fail
Quit the page

Figure 4.13 UML sequence diagram of creating account

4.2.3.2 Login

After creating an account, user can login to the system using ther username and password. Otherwise, user cannot enter the prediction page to get the information (Figure 4.8).

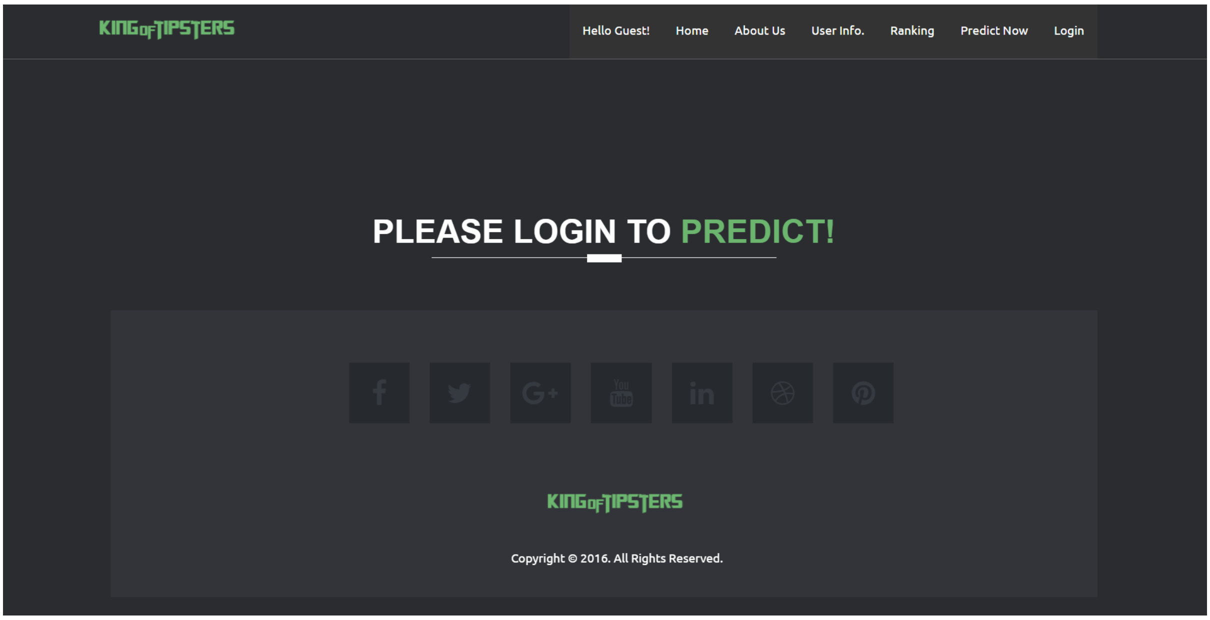


Figure 4.14 Predicting Page without Login

To login to the webpage, user can follow this way:

1. Click on the Login button on the menu bar

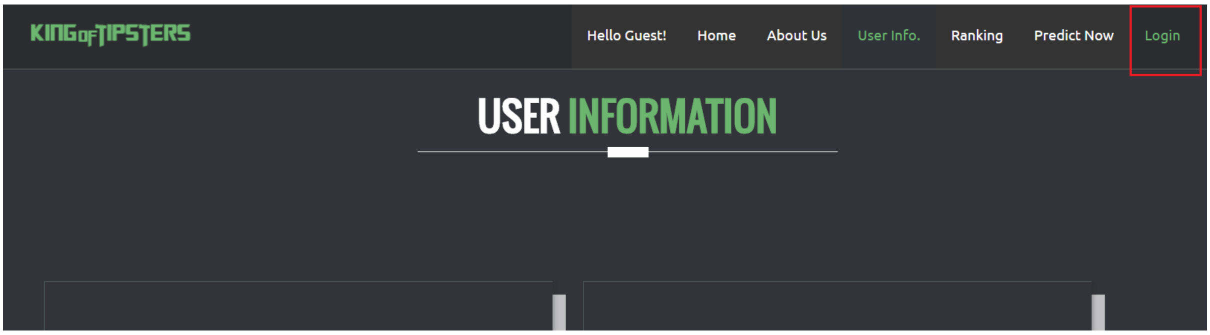


Figure 4.15 Login Button

2. Input the username and password
3. Click on the Login button

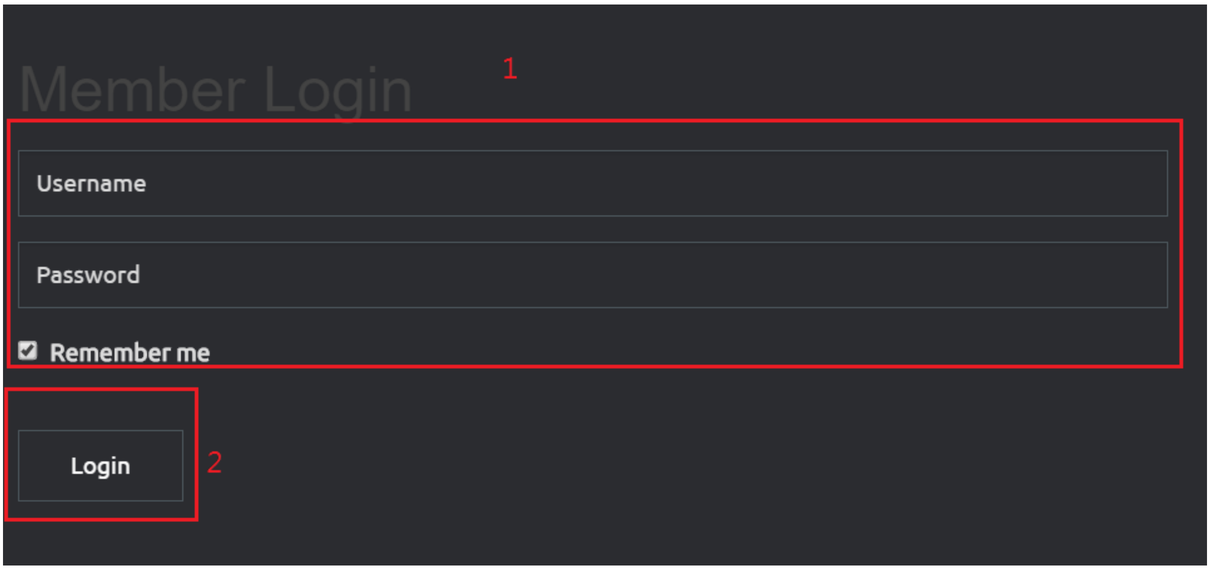
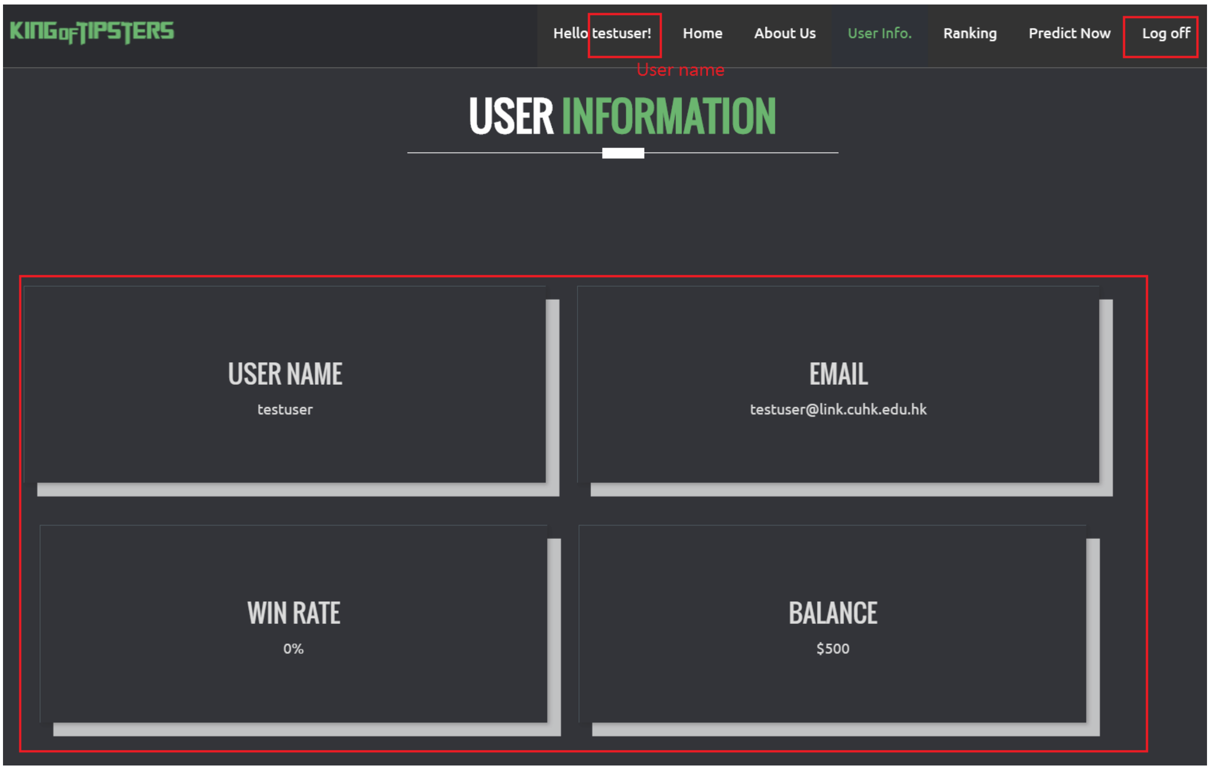


Figure 4.16 Login Area

4(a) If the login process is successful, the webpage will be back to the homepage. Users can also see their username on the menu bar, and the Login button will be replaced by the Log off button. They can also find out their user information, win rate and balance on the homepage.



USER NAME
testuser

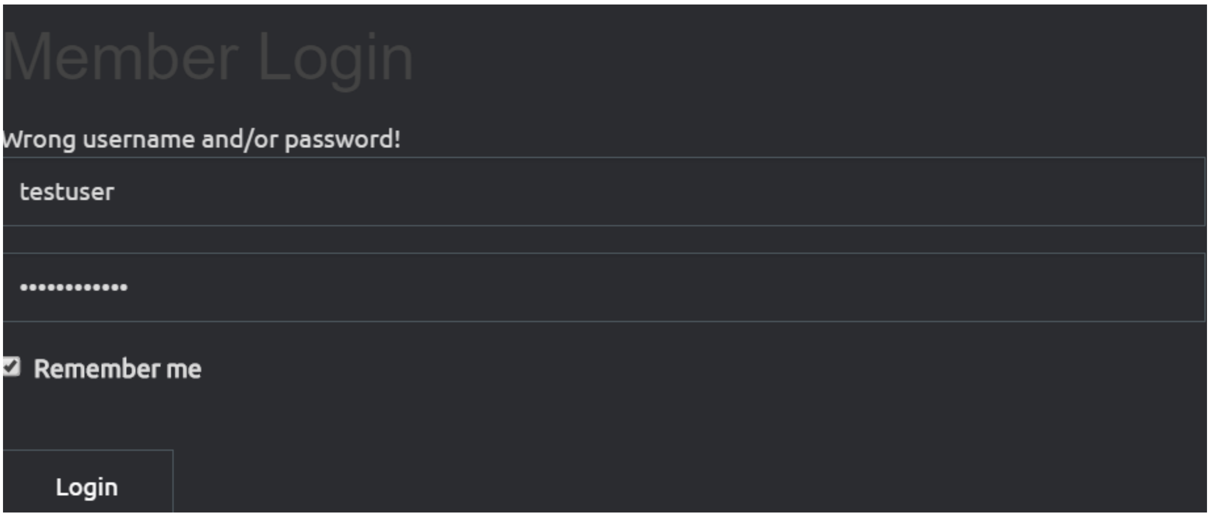
EMAIL
testuser@link.cuhk.edu.hk

WIN RATE
0%

BALANCE
$500

Figure 4.17 Homepage after Login

4(b) If user input their username or password wrongly, login will be unsuccessful and the error syntax will be displayed on the login area.



Member Login

Wrong username and/or password!

testuser

••••••••••••

☑ Remember me

Figure 4.18 Login Failed Demo

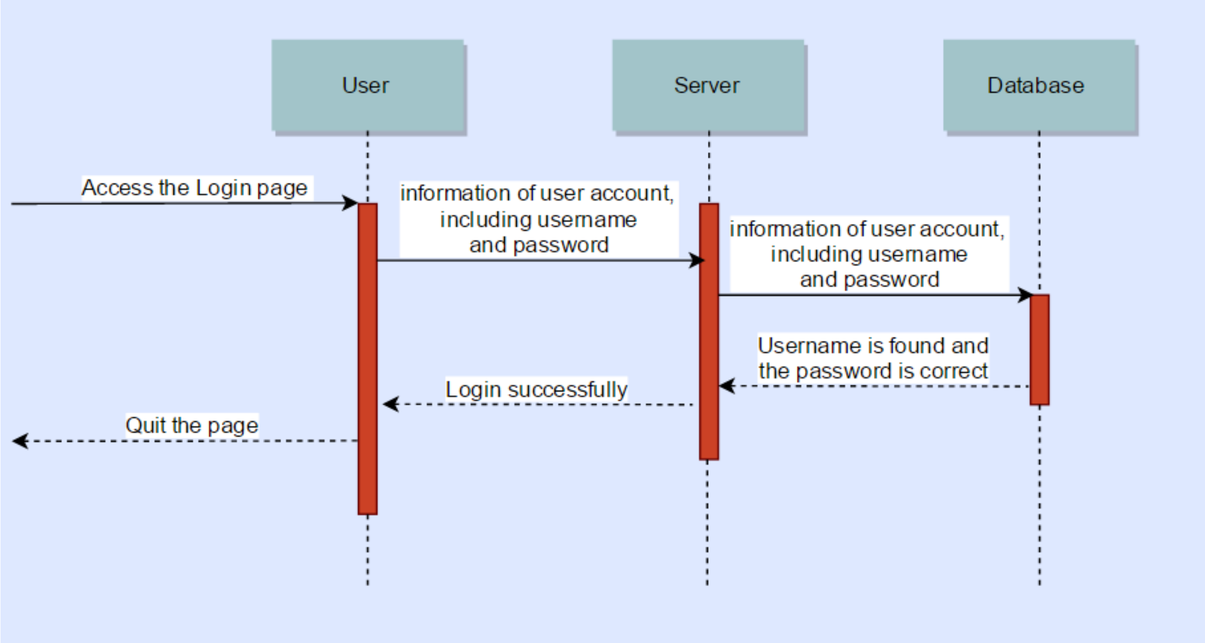

User
Server
Database
Access the Login page
information of user account, including username and password
information of user account, including username and password
Username is found and the password is correct
Login successfully
Quit the page

Figure 4.19 UML sequence diagram of Login

4.2.3.3 Prediction

When user have their own account and login to the system, they can access the prediction page and do the prediction action. To predict horse, user can follow this way:

1. Click on the Predict Now button on the menu bar

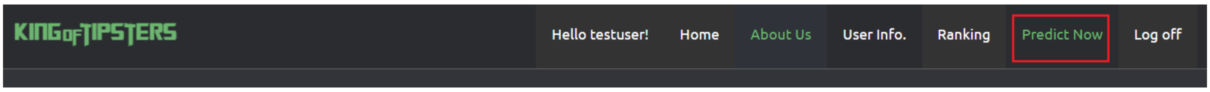
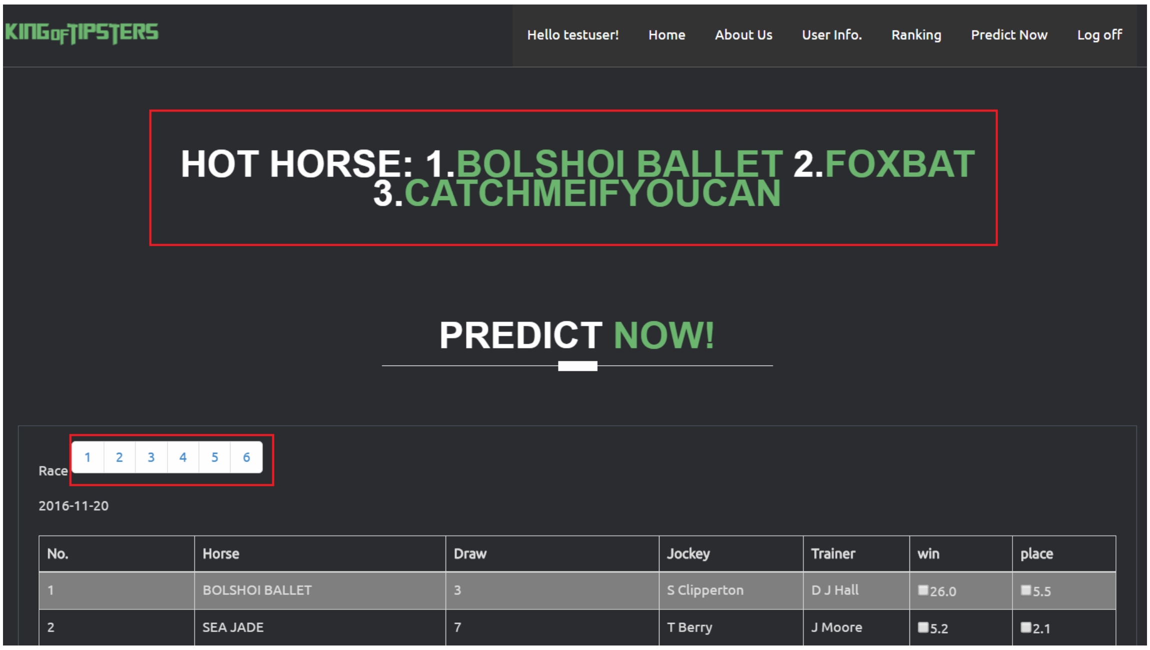


Figure 4.20 Predict Now button

2. User can see the three Hot Horses suggested by our algorithm on the prediction page. After that, user need to choose the race number that they want to predict.



HOT HORSE: 1.BOLSHOI BALLET 2.FOXBAT 3.CATCHMEIFYOUCAN

PREDICT NOW!

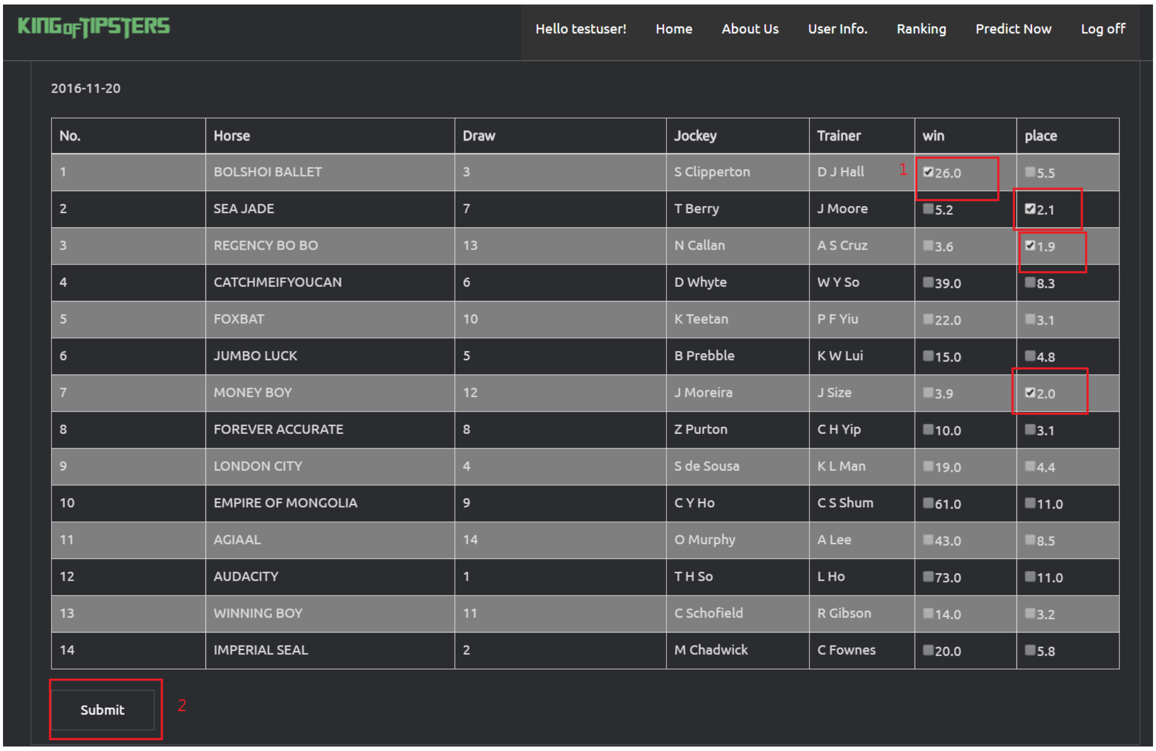
Race 1 2 3 4 5 6

2016-11-20

No.	Horse	Draw	Jockey	Trainer	win	place
1	BOLSHOI BALLET	3	S Clipperton	D J Hall	□26.0	□5.5
2	SEA JADE	7	T Berry	J Moore	□5.2	□2.1

Figure 4.21 Hot Horses and Race Number Button

3. The horse list of the race will be displayed below. User can pick at most one horse on the win column and three horses on the place column. The system does not allow user to tick more than the limit. Finally, click on the submit button.



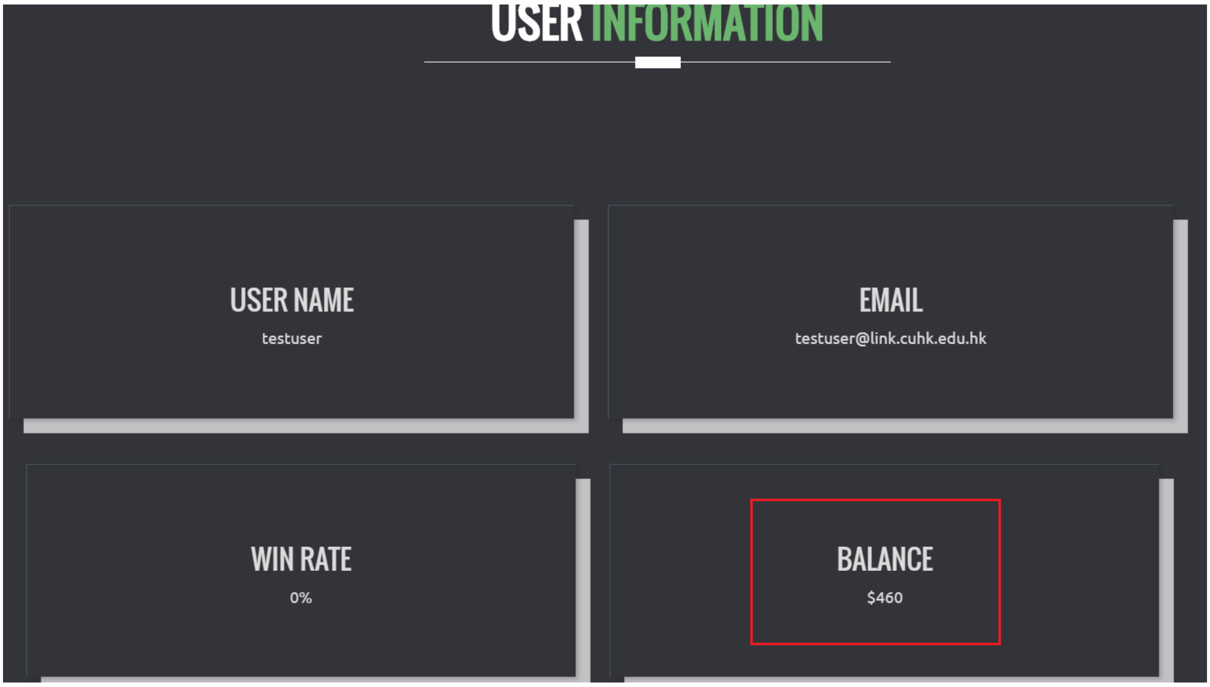
2016-11-20

No.	Horse	Draw	Jockey	Trainer	win	place
1	BOLSHOI BALLET	3	S Clipperton	D J Hall	☑26.0	□5.5
2	SEA JADE	7	T Berry	J Moore	□5.2	☑2.1
3	REGENCY BO BO	13	N Callan	A S Cruz	□3.6	☑1.9
4	CATCHMEIFYOUCAN	6	D Whyte	W Y So	□39.0	□8.3
5	FOXBAT	10	K Teetan	P F Yiu	□22.0	□3.1
6	JUMBO LUCK	5	B Prebble	K W Lui	□15.0	□4.8
7	MONEY BOY	12	J Moreira	J Size	□3.9	☑2.0
8	FOREVER ACCURATE	8	Z Purton	C H Yip	□10.0	□3.1
9	LONDON CITY	4	S de Sousa	K L Man	□19.0	□4.4
10	EMPIRE OF MONGOLIA	9	C Y Ho	C S Shum	□61.0	□11.0
11	AGIAAL	14	O Murphy	A Lee	□43.0	□8.5
12	AUDACITY	1	T H So	L Ho	□73.0	□11.0
13	WINNING BOY	11	C Schofield	R Gibson	□14.0	□3.2
14	IMPERIAL SEAL	2	M Chadwick	C Fownes	□20.0	□5.8

Submit

Figure 4.22 Prediction Demo

4. After the prediction is submitted, the system will display a syntax, ‘PREDICTED SUCCESSFULLY’. Besides, we have set an amount $10 for predicting each horse. The balance will be updated immediately in the homepage. During this demonstration, the user has picked four horses, so his balance is decreased from $500 to $460.



USER INFORMATION

USER NAME
testuser

EMAIL
testuser@link.cuhk.edu.hk

WIN RATE
0%

BALANCE
$460

Figure 4.23 Balance Updated

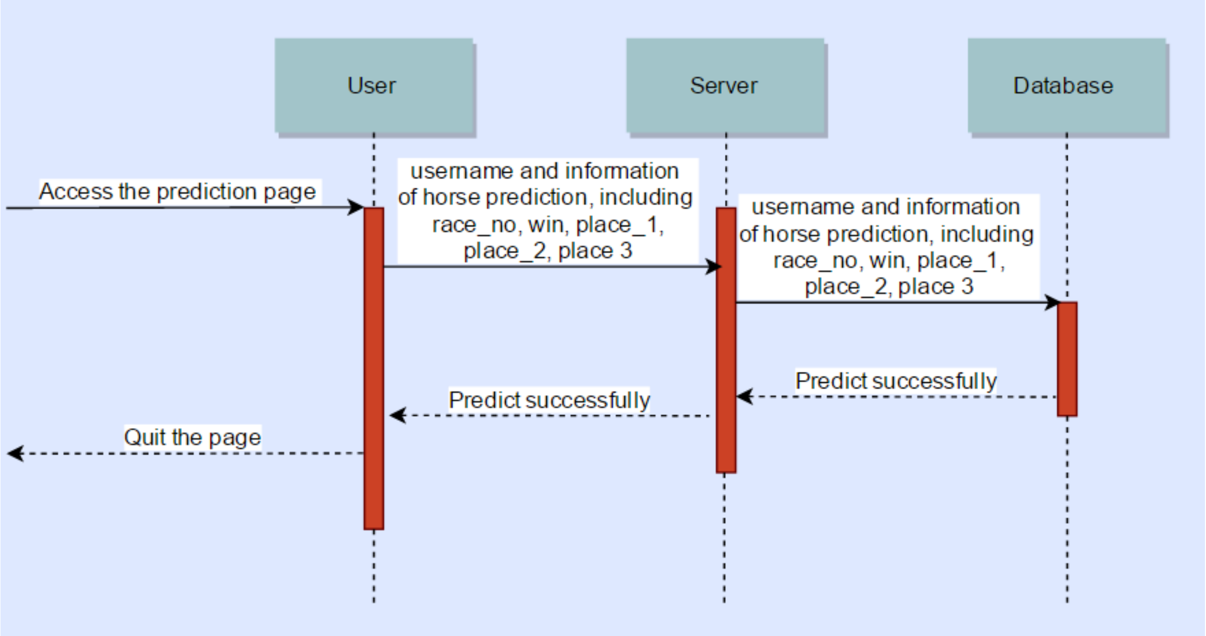

User
Server
Database
Access the prediction page
username and information of horse prediction, including race_no, win, place_1, place_2, place 3
username and information of horse prediction, including race_no, win, place_1, place_2, place 3
Predict successfully
Predict successfully
Quit the page

Figure 4.24 UML sequence diagram of prediction

5 Algorithm on giving suggestion

In this project, we developed this system, and using collective intelligence to give suggestion. In this section, we will talk about how we do a suggestion.

5.1 Rating a user

Actually the odds of each horse representing how many people buying it, if many people buy that horse, the odds of it would be lower. If we just buying the horse with lowest odds, it is a kind of collective intelligence. However, just follow the lowest odds to buy horse is not good because in the long run we would lose money. So, we need to do some filtering, so that we only collect the choices of people who has a high win rate.

5.1.1 Rate by win rate

We use win rate to determine a user is good or not. To get the win rate of users, we calculate it after the racing day, because we know the results already. If the user bought 10 races, and correctly predicted 5 of them, the win rate of the user is 50%, and the win rate is accumulative, next time if the user buys 5 races and all of them are correct, the win rate will become 66.7%.

At this moment, we decide to filter out the users whose win rate is below 40%(40% is adjustable, we will change it depending on the real situation). For example, in one particular horse race, there are 100 people did predictions on this match, we only consider those with win rate higher than 40%.

5.1.2 Rate by balance

Apart from win rate, balance is another factor that we are considering. It is because some users may prefer high risk and high reward. They may always try to predict horses with high odds and low win rate. So that they have a chance to win a large amount. However, their win rate should be relatively lower than those conservative users. But we should not classify those user be bad predictors. Since some user may earn a lot using the high risk strategy. So, we have set one more rule for those special cases. If a user own balance more than $500 (pre-assigned amount for every user), even if the win rate is below 40%, we also consider their prediction. Besides, we decide to give a penalty score for those conservative users when calculating. The detailed calculation will be shown in section 5.2.

5.1.3 Bonus rate for the high ranking user

We have mentioned in section 4.2 that our application provides a ranking system. The name of user with highest win rate, balance will be shown on the ‘Best Tipsters’ list. These users are the most accurate predictor among all users. Since we believe the prediction of those high ranking users have a high reference value, we propose to give a big bonus score for their predicted horses. The prediction of the best predictor will have the highest bonus score.

5.2 Suggestion of horses

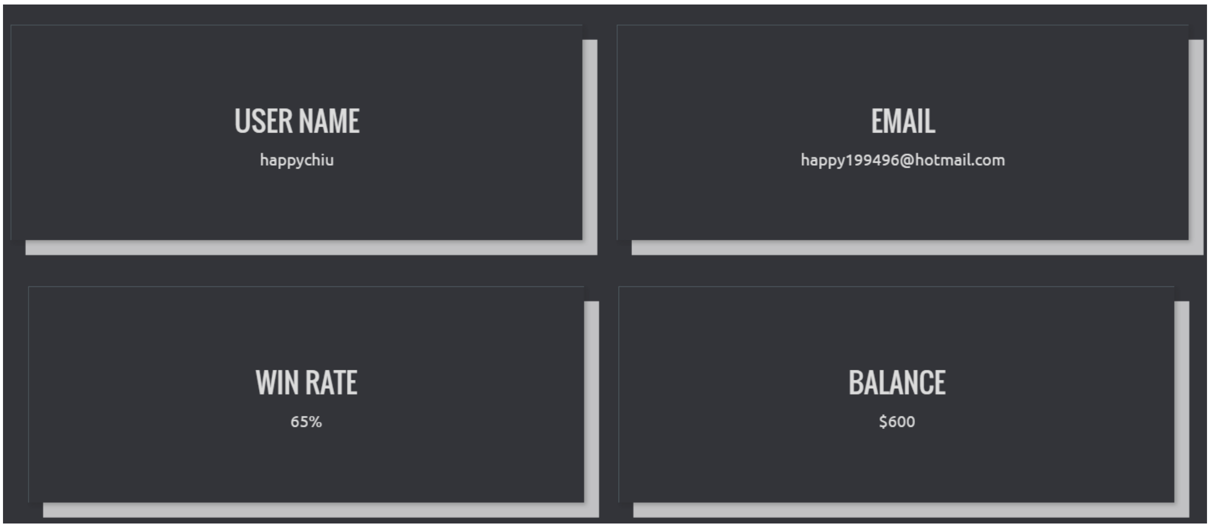
After we have selected those selected people, we need to do further to choose horses for suggestions. We look at the choices of each above users, and add scores to horses depending on users’ choices, and the score is calculated by the win rate and balance of that user. If the user is on the ‘Best Tipsters’, we will mutiply their scores for the chosen horses. The mutiplied value depends on the user’s rank. According the above rules, we obtain a formula:

$$Score = (WinRate + \frac{Balance - 500}{1000}) \times BonusForTopUsers$$

We are going to simulate some score calculations. Suppose there are 8 horses for users to choose. The initial score of the horses are 0.

Horse no.	1	2	3	4	5	6	7	8
Score	0	0	0	0	0	0	0	0

User No. 1:


USER NAME
happychiu

EMAIL
happy199496@hotmail.com

WIN RATE
65%

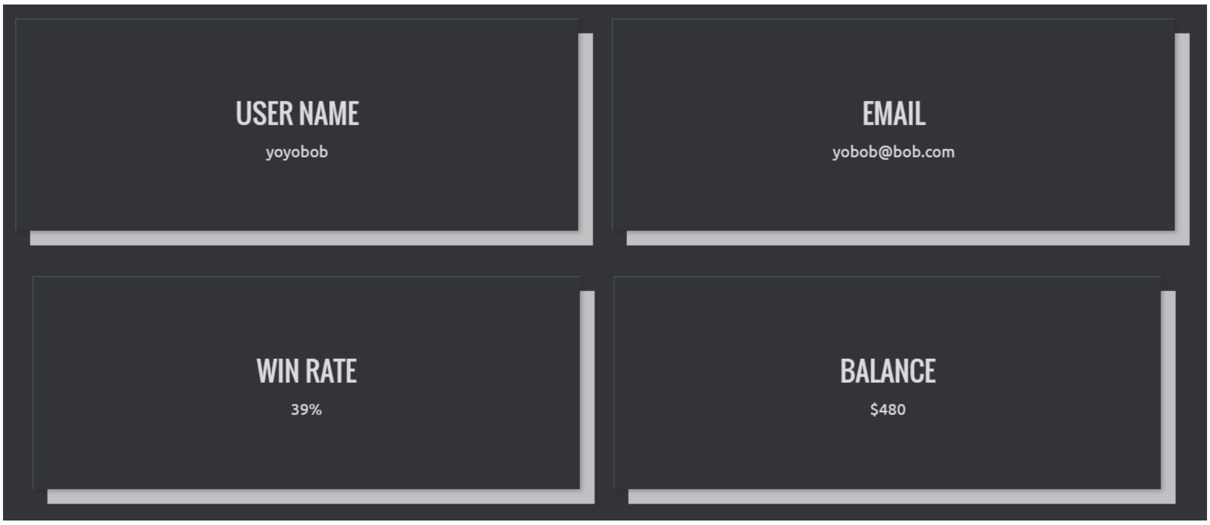
BALANCE
$600

Figure 5.1 User Information 1 for score calculation

This user predict horse 3 will win, and he choose horse 3, 5 for place. He is not in the list of 'Best Tipsters', so he cannot give the bonus score. The score he can give to each horses is $0.65 + \frac{600-500}{1000} = 0.75$.

Horse no.	1	2	3	4	5	6	7	8
Score	0	0	1.5	0	0.75	0	0	0

User No. 2:


USER NAME
yoyobob

EMAIL
yobob@bob.com

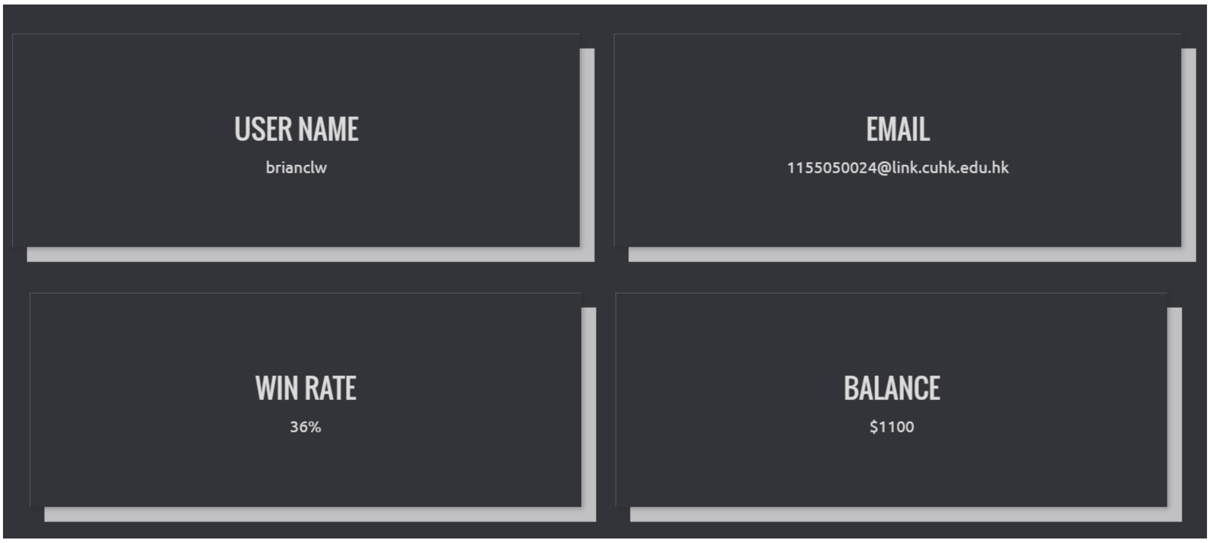
WIN RATE
39%

BALANCE
$480

Figure 5.2 User Information 2 for score calculation

This user has win rate lower than 40% and balance less than $500, although he can still do prediction, we do not calculate the choice of this user.

User No. 3:



USER NAME

brianclw

EMAIL

1155050024@link.cuhk.edu.hk

WIN RATE

36%

BALANCE

$1100

Figure 5.3 User Information 3 for score calculation

Although this user has win rate lower than 40%, his balance is higher than 500. So, we will use his prediction for calculating. Besides, he is the rank 3 ‘Most Money Made’ user, we give him a 18 bonus value for his prediction. This user predict horse 1 will win, and he choose horse 1, 5, 8 for place. The score he can give to each horses is

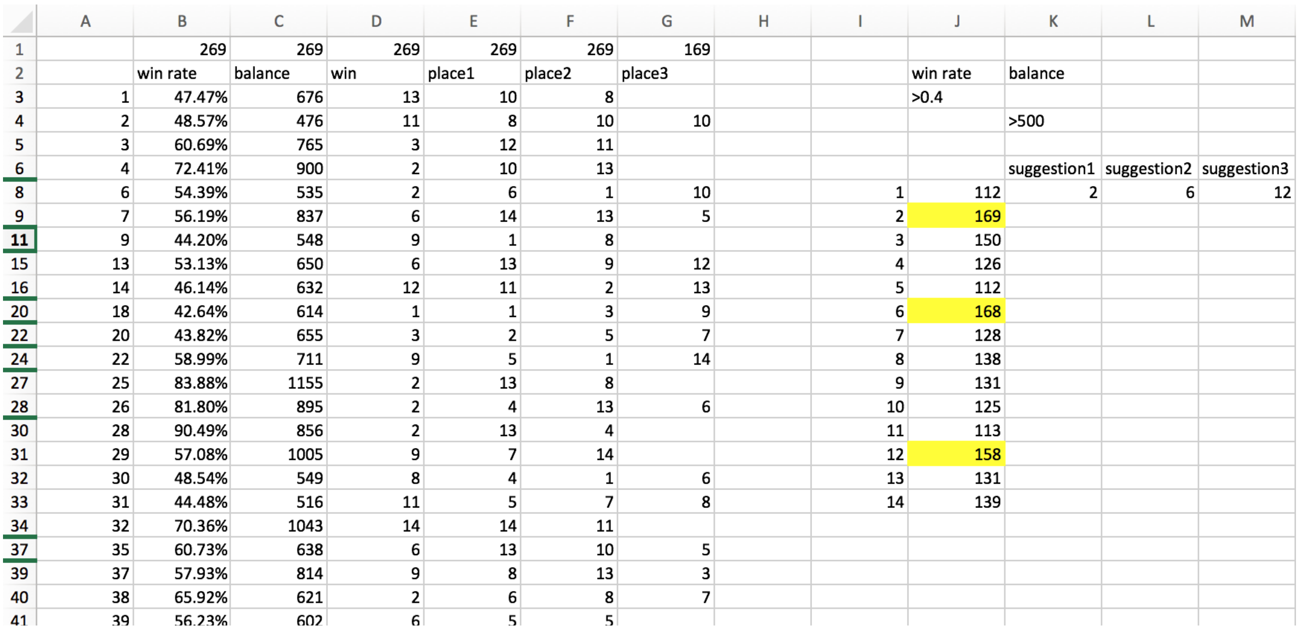
$$(0.36 + \frac{1100 - 500}{1000}) \times 18 = 17.28$$.

Horse no.	1	2	3	4	5	6	7	8
Score	34.56	0	1.5	0	18.03	0	0	17.28

After the predictions of these users, we find that horses 1, 5, 8 have the highest score. So, they will be the 3 hot horses for this match in this moment. The list of hot horses may change at any time after more predictions are submitted.

5.3 Simulation of buy records

At this moment, we don’t have real users, so we simulated a buy record of 500 users, and showed how we get the 3 suggestions.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1		269	269	269	269	269	169						
2		win rate	balance	win	place1	place2	place3			win rate	balance		
3	1	47.47%	676	13	10	8				>0.4			
4	2	48.57%	476	11	8	10	10				>500		
5	3	60.69%	765	3	12	11							
6	4	72.41%	900	2	10	13					suggestion1	suggestion2	suggestion3
8	6	54.39%	535	2	6	1	10		1	112	2	6	12
9	7	56.19%	837	6	14	13	5		2	169			
11	9	44.20%	548	9	1	8			3	150			
15	13	53.13%	650	6	13	9	12		4	126			
16	14	46.14%	632	12	11	2	13		5	112			
20	18	42.64%	614	1	1	3	9		6	168			
22	20	43.82%	655	3	2	5	7		7	128			
24	22	58.99%	711	9	5	1	14		8	138			
27	25	83.88%	1155	2	13	8			9	131			
28	26	81.80%	895	2	4	13	6		10	125			
30	28	90.49%	856	2	13	4			11	113			
31	29	57.08%	1005	9	7	14			12	158			
32	30	48.54%	549	8	4	1	6		13	131			
33	31	44.48%	516	11	5	7	8		14	139			
34	32	70.36%	1043	14	14	11							
37	35	60.73%	638	6	13	10	5						
39	37	57.93%	814	9	8	13	3						
40	38	65.92%	621	2	6	8	7						
41	39	56.23%	602	6	5	5							

Figure 5.3 Simulation of buy records

6 Limitation and difficulties

Number of users

In our research, we are doing collective intelligence, which is a type of crowd sourcing, so we need people to use our system. If we don’t have enough users, we can’t make a conclusion that collective intelligence on predicting horse racing is good or not. In many other successful crowd sourcing systems, there are thousands to millions people to contribute to that system. We think it is hard to achieve this amount of users, but we still try our best to promote our system, and hope that we can have more users.

7 Goals for Second Term

Design mobile application

We have finished the web page version of our system, we are going to implement it on mobile application, because we need to encourage people to use, and mobile version is more convenient for users. Also, while we developing mobile application, we can learn more skills, which is meaningful for us.

Improve the algorithm

After the mobile application is released, we can collect the predicted data of different users. Actually in our algorithm, there are many parameters that we asuumed, in the future we need to tune the parameters, so that we can get the best result.

On the other hand, there are many different methods to collect data, maybe not just filtering, we hope we can try more different models, trying is important in doing research, we hope that we can find the best way to do it.

8 Conclusion

In conclusion, our topic is Research in Collective Intelligence through Horse Racing in Hong Kong. In this semester, we have developed our website, which called King of Tipsters. We started our work in last summer, and keeo doing it steps by steps, it was a great experience for us to do a big project. The page is already connected to the server and database, and prepared to publish to public. Actually we are the first one to do collective intelligence on horse racing, and we are so excited to see the outcome.

In these few months, we really leared a lot of things. Before doing this project, although we have learned the basic knowledge of database, web development, and some knowledge about server, we didn’t have enough solid working experience on them. We revised a lot and did many researches to learn how to implement them to our system, we effort has paid off. For example, we learned how to connect our server to database, we know how to write PHP code to perform some actions, learning new things is always the best in studies.

In the next semester, we will keep on doing on this topic and pay more effort on it, we hope that we can develop our system to mobile application, although the way to our goal is challenging, we will do our best to accomplish it.

9. Acknowledgements

The completion of this undertaking could not have been possible without the participation and assistance of so many people. Their contribution are sincerely appreciated and gratefully acknowleged. We would like to express our deep appreciation and indebtedness to the following:

Dr. Michael R. Lyu, Mr. Yau for their endless support, kind and leading us during the project.

LYU1603 team for their spiritually support, thank you.

References

Condorcet, J.-A.-N. de Caritat, marquis de. "Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix." Paris: Imprimerie Royale, 1785. (http://mathworld.wolfram.com/CondorcetsJuryTheorem.html)

Doan, A., Ramakhrisnan, R. and Halevy, A. (2011) Crowdsourcing Systems on the World-wide Web. Communications of the ACM, 54, 86-96.

Estellés-Arolas, E. and González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information Science, 38, 189.

Howe, J. (2006) The rise of crowdsourcing. Wired Magazine. June 2006.

Krause, J., Ruxton, G. and Krause, S. (2009) Swarm Intelligence in Animals and Humans. Trends in Ecology and Evolution, 25, 28-34.

Massanari, A. (2012) DIY design: How crowdsourcing sites are challenging traditional graphic design practice. First Monday, 17.

Salminen, J. (2013a) Collective intelligence on a crowdsourcing site. GBI Working Paper. Available at https://sites.google.com/site/gbialternative1/working-papers.

Salminen, J. (2013b) Crowdsourcing innovations as a search process. PIN-C 2013, Lahti, Finland. 18.-20. June 2013.

Zhao, Y. and Zhu, Q. (2012) Evaluation on crowdsourcing research: Current status and future direction. Inf Syst Front.