

# BetSmart

## MIS510 Final Report



Cai Chen, Ximing Yu, Ying Jin  
Spring 2010

## Table of Contents

1.	Introduction to BetSmart .....	3
2.	Site Tour .....	3
2.1	General soccer fans.....	3
2.2	Online Better.....	6
3.	Business Model .....	8
3.1	Revenue Model .....	8
3.2	Competitive Analysis.....	8
4.	System Architecture.....	9
4.1	Google App Engine.....	10
4.1.1	Advantages of GAE.....	11
4.1.2	Disadvantages of GAE .....	11
4.1.3	Suggestions to using GAE.....	11
4.2	Data Collection.....	12
4.2.1	Gazzetta dello Sport.....	12
4.2.2	BetExplorer .....	12
4.2.3	Football-Italia .....	12
4.3	API Integration .....	12
5.	Data Mining Component.....	13
5.1	Preparation for data mining.....	13
5.2	Data mining using Weka .....	14
5.3	Evaluation of the Prediction Model .....	14
6.	Conclusion.....	15
7.	Appendix .....	15
	Team Contribution.....	15

## 1. Introduction to BetSmart

Soccer is the most popular sport in the world. The performance of teams and players can emotionalize their fans greatly. As a byproduct, soccer betting is turning out to be a very profitable field. BetSmart focuses on guiding betting in Serie A of Italy, one of the top soccer leagues in the world. The website provides all the information needed for doing betting, including news, club and player information, matches statistics and predictions of coming matches.

## 2. Site Tour

BetSmart targets at different segments of users: the first segment would be general soccer fans as they can browse latest news, club standings, match results, videos and so on. The second segment would be online betters. It provides prediction of the coming matches based on history data and data mining techniques. Also it displays odds provided by all betting companies worldwide and labels the highest odds for the 3 outcome (home-win, draw and away-win). Additionally, in case experienced betters are interested in statistical data like odds history and history match results, based on which they can make new predictions, the website collects history data for the past 12 years.

### 2.1 General soccer fans

There are five main categories of information in this website: News, Matches, Standing, Clubs and Statistics (Figure 1).

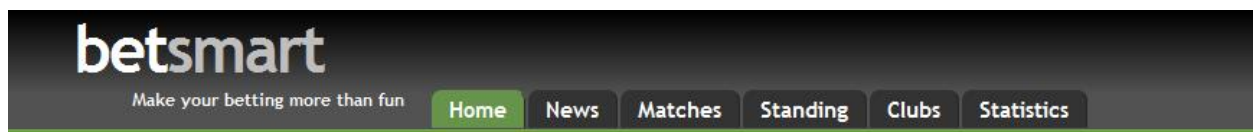


Figure 1 Navigation of the website

In the home page, the website provides latest news and current Standing of Serie A (Figure 2).



**Latest News**

- Moggi: "Swiss Cards? Never date" ⌚ May 04, 2010
- [Meaning, hard to replicate Lotito "The derby? Latium Provocations"](#) ⌚ May 04, 2010
- Outside, Blue Sky HD channel and a World Cup attack ⌚ May 04, 2010
- Camoranesi, Juventus goodbye "ended cycle" ⌚ May 04, 2010
- Mourinho Against the Senses "Inter Demands Respect" ⌚ May 04, 2010
- Galliani: "Pato will not be shaken" and "Duck" trains ⌚ May 04, 2010
- Cruyff: "That whole team Mourinho But I do not like" ⌚ May 04, 2010
- The Office of Rainier "There are no favorites" ⌚ May 04, 2010
- Lippi, with Totti Discretion Which you call? Vote ⌚ May 04, 2010
- Buffon at Juventus and the future "If England will go away" ⌚ May 03, 2010

[Read more](#)

**Serie A Standing**

Club	W	D	L	Pts
Inter	22	10	4	76
Roma	22	8	6	74
Milan	19	10	7	67
Sampdoria	18	9	9	63
Palermo	17	10	9	61
Napoli	14	14	8	56
Juventus	16	7	13	55
Genoa	13	9	14	48
Fiorentina	13	7	16	46
Bari	12	10	14	46
Parma	12	10	14	46
Chievo	12	8	16	44

Figure 2 Latest news and current Serie A standing in home page

By clicking on the news title, users will be directed to the original page containing that piece of news. For the convenience of users, the news pages are all translated into English by Google Translation.

At the bottom of home page (Figure 3) are introduction to the website, information source and site map. The website also provides some small but functions, including a quick search box and an “AddThis” button through which users can share this website on their Facebook or Twitter.



Figure 3 Utilities at the bottom of home page

In “News” and “Standing” page, more detailed news and standing information are provided than in the home page.

In the “Matches” page, all matches in this season are listed by weeks. Users can also use the “Go to week” tool to go directly navigate to a particular week. When clicking on the score of one match, users can get the detailed information about that match. The information includes video of the match from Youtube, detailed report about the match, and so on (Figure 4).



Figure 4 Video about the match from Youtube

On the Clubs page, all the 20 clubs in serie A are listed. By clicking on the logo, users can go to the detail page about that club. Information such as leadership, technical stuff, honors, squad and matches in this season is listed (Figure 5).

### Milan



**General Information**

- Head Office: Via Turati 3 - 20121 Milano Tel. 0262281 - Fax 026598876
- Training Camp: Centro Sportivo Milanello - Via Milanello 25 Carnago (VA) - Tel. 0331991174
- Foundation Year: 1899
- Team Colors: Jersey striped red-black, white shorts with red stripes-black socks with red border
- Stadium: GIUSEPPE MEAZZA

**Leadership**

**Technical Staff**

**Honors**

**Squad**

Num	Name	Birthday	Pos
1	Nelson De Jesus Silva Dida	1973-10-07	G
12	Christian Abbiati	1977-07-08	G


### Milan Season Matches

W	Home	Away
1	Siena	1 - 2 <i>Milan</i>
2	<i>Milan</i>	0 - 4 Inter
3	Livorno	0 - 0 <i>Milan</i>
4	<i>Milan</i>	1 - 0 Bologna
5	Udinese	1 - 0 <i>Milan</i>
6	<i>Milan</i>	0 - 0 Bari
7	Atalanta	1 - 1 <i>Milan</i>
8	<i>Milan</i>	2 - 1 Roma
9	Chievo	1 - 2 <i>Milan</i>
10	Napoli	2 - 2 <i>Milan</i>
11	<i>Milan</i>	2 - 0 Parma
12	Lazio	1 - 2 <i>Milan</i>
13	<i>Milan</i>	4 - 3 Cagliari
14	Catania	0 - 2 <i>Milan</i>
15	<i>Milan</i>	3 - 0 Sampdoria

Figure 5 Detail information about a club

Users can also check out the information about one particular player by clicking on the name in the squad list. Photos from Flickr are provided, as well as the player's information, performance in this season and career paths (Figure 6).

### Filippo Inzaghi








**FILIPPO INZAGHI**

Club	Milan
Born place	Piacenza
Birth date	1973-08-09
Height	181cm
Weight	74kg
Nationality	Italia
Position	Striker
Jersey number	9
Serie A debut	1995-08-27

#### Serie A Season 2009-10

Appearance	23
Goals	2
Minutes played	812
Yellow cards	4
Red cards	0

#### Player Photos

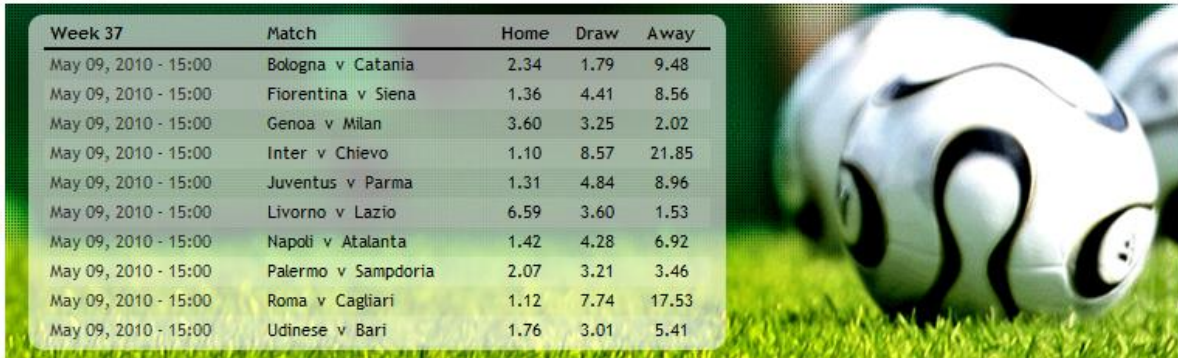






Show more...

Figure 6 Detail information about a player

## 2.2 Online Better

There is a list of matches in the coming week on the home page (Figure 7). By clicking on the match, users can easily go to the page containing the statistics information about that match.



Week 37	Match	Home	Draw	Away
May 09, 2010 - 15:00	Bologna v Catania	2.34	1.79	9.48
May 09, 2010 - 15:00	Fiorentina v Siena	1.36	4.41	8.56
May 09, 2010 - 15:00	Genoa v Milan	3.60	3.25	2.02
May 09, 2010 - 15:00	Inter v Chievo	1.10	8.57	21.85
May 09, 2010 - 15:00	Juventus v Parma	1.31	4.84	8.96
May 09, 2010 - 15:00	Livorno v Lazio	6.59	3.60	1.53
May 09, 2010 - 15:00	Napoli v Atalanta	1.42	4.28	6.92
May 09, 2010 - 15:00	Palermo v Sampdoria	2.07	3.21	3.46
May 09, 2010 - 15:00	Roma v Cagliari	1.12	7.74	17.53
May 09, 2010 - 15:00	Udinese v Bari	1.76	3.01	5.41

Figure 7 List of coming matches

The website provides its own prediction about the result (Figure 8).

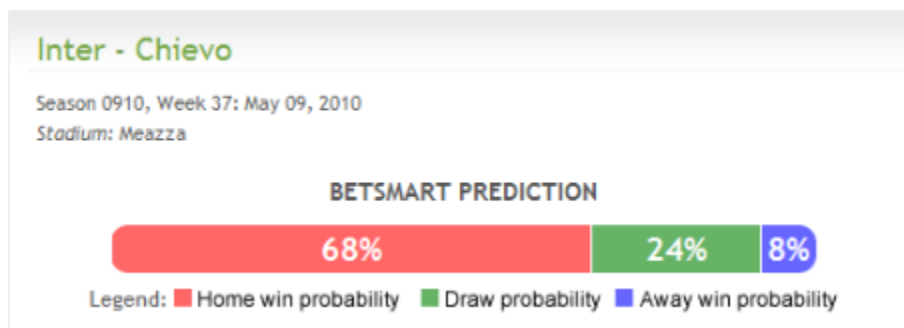


Figure 8 Prediction result from the website

Season ranking trend is provided to show the long term performances of the two teams (Figure 9 Season ranking trend). And recent matches result is provided to show the short term performances (Figure 10).



Figure 9 Season ranking trend

### RECENT MATCHES

W	Inter	W	Chievo
36	Lazio 0 : 2 Inter	36	Chievo 1 : 2 Napoli
35	Inter 3 : 1 Atalanta	35	Fiorentina 0 : 2 Chievo
34	Inter 2 : 0 Juventus	34	Chievo 2 : 0 Livorno
33	Fiorentina 2 : 2 Inter	33	Palermo 3 : 1 Chievo
32	Inter 3 : 0 Bologna	32	Chievo 1 : 2 Sampdoria

Figure 10 Recent matches of teams

Squad information is also provided for users to check the formation and players information (Figure 11). The website also collects rival history of the two teams since season 98-99 for users (Figure 12). These are all very important for an online better to do the analysis and final decision.

### SQUAD

Inter (Home)

Predicted formation (4-2-3-1)		
12-Julio Cesar/ 13-Maicon, 2-Cordoba, 25-Samuel, 4-Zanetti/ 5-Stankovic, 19-Cambiasso/ 9-Eto'o, 10-Sneijder, 45-Balotelli/ 22-Milito		
Substitutes	Injured	Suspended
1-Toldo; 23-Materazzi; 26-Chivu; 8-Motta; 11-Muntari; 17-Mariga; 27-Pandev;		None;

Figure 11 Squad information

### RIVAL HISTORY (SINCE SEASON 98-99)

Season	Match Result		
01-02	Chievo	2 : 2	Inter
01-02	Inter	1 : 2	Chievo
02-03	Chievo	2 : 1	Inter
02-03	Inter	2 : 1	Chievo
03-04	Chievo	0 : 2	Inter
03-04	Inter	0 : 0	Chievo
04-05	Chievo	2 : 2	Inter

Figure 12 Rival history

After looking through the information provided by the website, users should be ready to do the betting. Users always want to know the highest odds for each result so that they can gain more. In order to help users save time searching in different websites, BetSmart has collected odds from different betting companies and highlighted the highest odd for each result (Figure 13). Users can simply see the highest odd and do online betting on that company by clicking on the company name.

**Match Odds**

Below are odds provided by all betting companies for the match. Click on corresponding links below to access their betting pages.

Bet Provider	H	D	A
Bet-At-Home	1.1	8.0	17.0
Bet365	1.13	7.0	15.0
Bet770	1.1	7.75	18.0
Betako	1.12	7.2	18.11
BetChronicl	1.11	7.6	30.0
BetClic	1.12	8.0	18.0
BETDAQ exch	1.11	9.36	31.4
BetKlass	1.13	8.55	20.0
Betliner ex	1.11	9.55	31.71
Betsson	1.08	9.0	25.0
BETSSON exc	1.11	9.36	28.55

Figure 13 Match odds

## 3. Business Model

### 3.1 Revenue Model

Betsmart is a place where you can have fun as well as make profit. Therefore, it is expected to be attractive to general users and profit-seekers. Here are several sources that we can make revenue from:

- Betting companies: As users who want to bet on this website will be directly linked to the betting companies, this is definitely bringing in new customers to those betting companies. Plus, when it recommends the best odds for betters, the betters are very likely to choose the



recommended betting companies. Considering all these, the betting companies would be happy to partnership with Betsmart.

- Advertisements: Because this website is focused on soccer game and soccer betting, the users are very likely to be soccer fans and online betters. Therefore, it would be more effective to post soccer-related ads in this website than in others.
- Although for now the website is providing the prediction function freely, it can later set up membership mechanism which means that only members are open to the most critical content such as the recommendation and prediction. This way, membership charge could be another revenue source.

### 3.2 Competitive Analysis

Betsmart is a one-stop site for soccer fans. It provides rich information in multiple formats: reports, statistics, pictures, videos and so on. Our major competitors defined are: Italian Football, La Gazzetta dello sport and BetExplorer.

Italian Football: It provides rich information on matches and the history match results. But it does not contain anything related to betting. Besides, it doesn't provide any videos of the matches. Basically, it is a good information source, but not versatile enough.

La Gazzetta dello: It has very good information of the clubs and players in Italian Serie A. Nevertheless, because it is a newspaper site, it focuses solely on providing information and facts.

BetExplorer: As indicated by its name, it is a professional website for betting as it has rich statistical history data. But it does not include any other general information that might interest the soccer lovers. The biggest weakness of this website is that it does not analyze the history data, let alone making predictions.

## 4. System Architecture

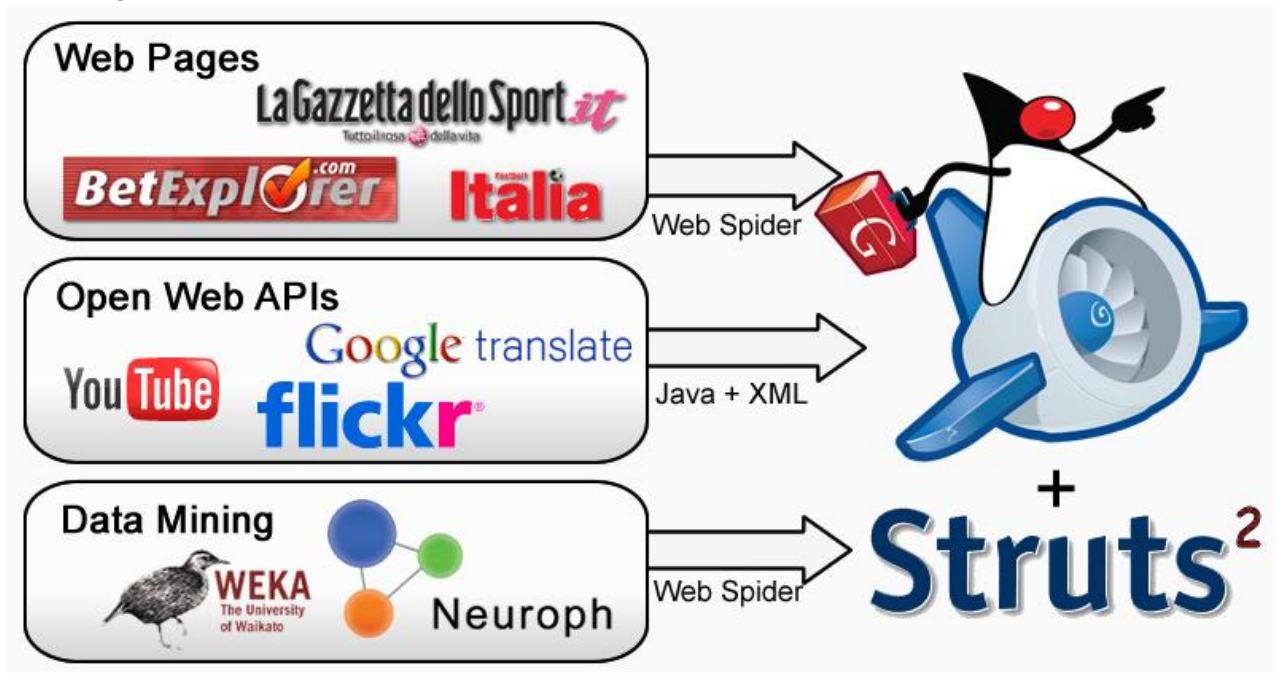


Figure 14 System Architecture

Figure 14 displays the architecture of our system. The whole system is built on Google App Engine and we choose Struts 2 as the web framework. We collected our data mainly from 3 soccer-related websites: Gazzetta dello Sport, BetExplorer and Football-Italia. Intelligent spiders are built to minimize the efforts needed to collect all the information. Open web APIs, such as YouTube, flickr, Google Translate and Google Visualization, are used to provide multimedia and multilingual components to users. To realize match predication, we utilized the Logistic Regression provided by Weka. Following is a guideline for our system architecture part.

Section 4.1 introduces what Google App Engine is and the cost-benefit of using Google App Engine. Section 4.2 describes our sources of data and how to build the spiders for each data source. Section 4.3 briefly covers how to integrate each web API into our system. The data mining component will be discussed with detail in Section 4.

#### 4.1 Google App Engine

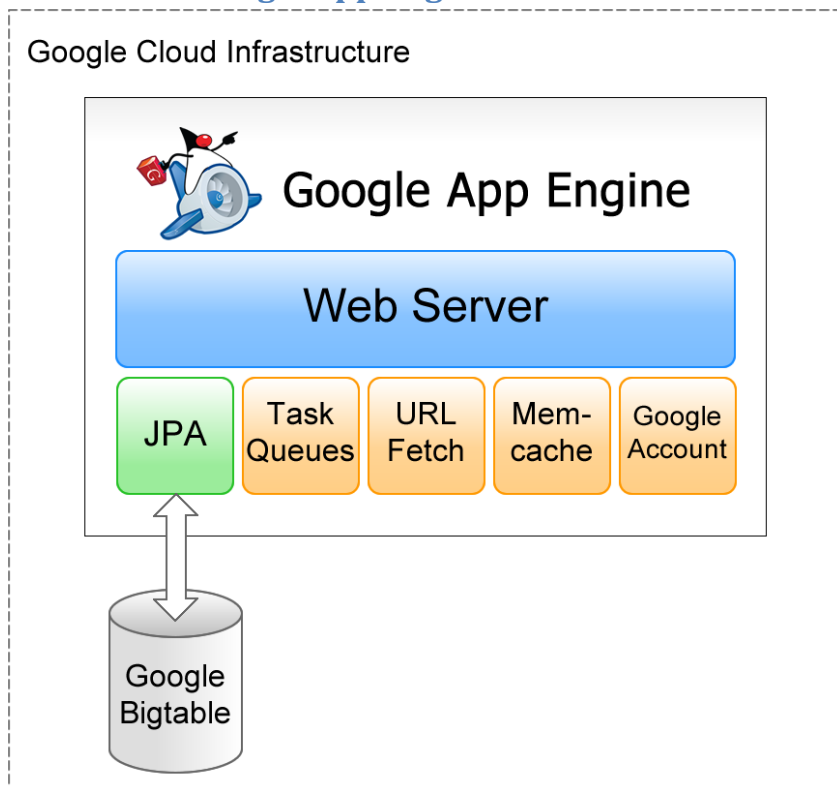


Figure 15 Architecture of Google App Engine

Google App Engine (GAE) lies in the cloud of Google and it lets users run their web application on Google's infrastructure for free when the application is under a certain quota. The current quota is 500 MB of storage and 5 million page views a month, which satisfies the needs of our proposed project.

Currently GAE supports two runtime environments, Java and Python. We chose to Java version since members of our group are more familiar with Java. The architecture of Java version GAE is shown in Figure 15. At the core of GAE is a web server, complemented by data access facilities and various other services enabled by Google.

**JPA** (Java Persistence API) specifies the standards for mapping object to tables, irrelevant of the implementation of the relational database. The data storage behind GAE is Google Bigtable, which is

more like a multidimensional map than relational database. This makes the implementation of JPA in GAE a little bit different from common ones. The JPA implementation of GAE is based on DataNucleus.

**Task queues** allow requests to be queued and executed in parallel. Since GAE has a 30-second limit on the execution of time of each request, task queues help a lot when sophisticated transactions are needed. With task queues, developers can divide a big task into several smaller ones and put all them into the queue and GAE will manage that queue for users.

**URL fetch** enables Java programs to create URL objects which can be used to exact contents of other URLs. The spiders for collecting data are based on URL fetch and we will discuss how to build the spiders in Section 4.2.

**Memcache** provides an implementation of Java L2 Cache and the benefits of using Memcache is that developers can store the commonly used data in Memcache (memory) and thus reduce the number of times for data retrieval from data store. In the developing process, we find that using Memcache can tremendously boost the responsiveness of our application.

**Google accounts** can also be integrated into GAE so that the application can authorize users based on their Google accounts. For our application, we use Google accounts to specify the system administrator of the site so that private pages can be properly protected.

#### 4.1.1 Advantages of GAE

Based on our experience developing with GAE, we found GAE to be of both advantages and disadvantages in different aspects. The advantages in using GAE are:

- GAE is *free of charge* for small or mid-scale application. It saves our time and money to find a scalable and stable server for hosting the application.
- The project structure of GAE is similar to common web project, so the learning curve is not very steep for those with web development experience on web servers like Tomcat.
- Google provides an Eclipse plugin for developing GAE. The plugin contains an implementation of web server (Jetty) and data store which can run on local developer machines. The debugging tool with the plugin is very useful.
- GAE provides a well-configured JPA implementation and the only codes needed for object persistence is the Java annotations on each entity class.
- GAE provides a lot of handy services, such as Task Queues, URL Fetch, Memcache, etc. GAE is shipped with default configurations for most of the services, so minimum configuration is required.

#### 4.1.2 Disadvantages of GAE

The disadvantages with GAE include:

- There are many restrictions on the Java runtime of GAE. For example, the program is not allowed to write files to the server, though the program can read from files transferred to the server. Library for parsing XMLs, like jdom, cannot run on GAE, which results in a lot of pains when processing XML retrieved from other web services.
- When the application codes are uploaded to Google's servers, the application may behave differently from running on local machines. The reason for this is that the Google's server has more restrictions, such as 30 seconds limit per request, 1 MB limit per object cached, etc. As there is no way to do debug on Google's server, these discrepancies are very annoying.
- Querying data from the data store is extremely slow on Google's server. As a result, the application is intolerably slow if we do not use Memcache to store often used data.

### 4.1.3 Suggestions to using GAE

Here we would like to provide some suggestions to future teams that may consider using GAE for their web mining project.

- Make the data query as simple as possible. Key based object retrieval is most efficient with GAE. Using sophisticated SQL-style query will often result in errors.
- If a web page need to load multiple types of data, they should split the page request to several phases and store intermediate data in Memcache, or the request process time will most likely exceed the 30 seconds limit.
- For doing data mining, we suggest exporting all the data from GAE to locally installed RDBMS, like SQL Server. Traditional relational DBMS fits better for data mining task of the required scale in this class.

## 4.2 Data Collection

The data of our project are collected from 4 websites, which are described by source as below.

### 4.2.1 Gazzetta dello Sport

We collected mainly club and player information from this website (<http://www.gazzetta.it>). Our spider follows a pattern and goes to each club's page, which contains the club information and a link to the club squad page. The spider parses the club info and retrieves the contents in the squad page, which are a list of URLs to each player's profile page. The spider collects the URLs of all the player profile pages and wraps them into task queues for future processing. When the task for parsing one player profile page is processed, a new spider is initialized to go into that player profile page and extract all the player information.

Additionally, we created a spider that collects probable formation for each club. The spider goes to a certain URL for each club, which contains the probable formation in XML format. Parsed data are then stored.

Gazzetta dello Sport provides news feeds of latest Serie A news for users to subscribe. We built an RSS feed parser to collect those news.

### 4.2.2 BetExplorer

BetExplorer (<http://www.betexplorer.com>) contains the betting history and match results of all matches since season 1998~1999. Our spider goes to each season's page and extracts all the match results and corresponding odds.

### 4.2.3 Football-Italia

Football-Italia (<http://football-italia.net/>) provides detailed match reports since season 2004~2005. Our spider first goes to a page listing all 38 weeks of this season (2009~2010). On going into the page for each week, the spider collects the URLs for web pages to match reports. We reassemble the URLs so as to collect URLs to matches between the same two teams in other seasons (from season 2004~2005 to season 2008~2009). Hereby we collected match reports for all the relevant matches through seasons. Scores, goal scorers, line-ups and booking information are also collected alongside the match reports.

## 4.3 API Integration

To improve user experience, we incorporated several open web APIs.

## YouTube

YouTube video URLs are collected for each match in this season. When using the YouTube API, we specified the season number, week number, home/away team name and home/away team scores as search criteria for retrieving the videos. The results are ranked by relevance and the first item on result list is stored for each match. By this way, we are able to provide satisfactory video highlight for each match.

## Flickr

flickr is used to incorporate pictures for players so that fans can see the player's pictures in addition to the profile image.

## Google Translation

We used a Java-based API (<http://code.google.com/p/google-api-translate-java/>) that bridges Google Translation and Java. When parsing the Gazzetta dello Sport news feeds, we call this API to directly translate all the news captions and summaries from Italian into English, since our targeted users are all English-readers.

## Google Visualization API

Google Visualization API is used in the project to provide users with a ranking trend graph of two opponent teams. By looking at the trend in the graph, people can easily tell whether a team is doing good or bad in either short term or long term.

# 5. Data Mining Component

The data mining component of the website aims in providing prediction result of a match.

## 5.1 Preparation for data mining

The training dataset generated for data mining consists of 3484 instances, each of which has 24 attributes. 23 of them are used as input variables and the attribute result is output variable, with 3 different class labels which are homewin, draw and awaywin.

The attributes of one instance cover both long term and short term performances of the two teams (Table 1).

Feature	Description	Feature	Description	Feature	Description
week	Time when the match took place	Home Points 4week	Points earned in the last 4 weeks (home team)	Home Points 10week	Points earned in the last 10 weeks (home team)
Home Points 30week	Points earned in the last 30 weeks (home team)	Away Points 4week	Points earned in the last 4 weeks (away team)	Away Points 10week	Points earned in the last 10 weeks (away team)
Away Points 30week	Points earned in the last 30 weeks (away team)	Hometeam Homeaway winratio	Home Team Ratio of Home wins to Away Wins	Awayteam homeaway winratio	Away Team Ratio of Home wins to Away Wins
Home drawratio	Home Team Ratio of Draws to Total Matches	Awaydraw ratio	Away Team Ratio of Draws to Total Matches	Homeavg Goalfor 10 week	Home Team Average goals For in last 10 matches
Homeavg	Home Team	Awayavg	Away Team	Awayavg	Away Team Average

goalfor 40 week	Average goals For in last 40 matches	Goalfor 10 week	Average goals For in last 10 matches	goalfor 40 week	goals For in last 40 matches
Hometea mhome winratio	Home Team Percentage of Home games won in one year	Awayteam awaywin ratio	Away Team Percentage of Away games won in one year	Hometeam home drawratio	Home Team Home Draw Percentage
Awaytea maway drawratio	Away Team Away Draw Percentage	Homeavgg oalagaint 10week	Home Team Average goals in last 10 weeks	Homeavgg oalagaint 40week	Home Team Average goals in last 40 weeks
awayavgg oalagains t10week	Away Team Average goals in last 10 weeks	Awayavgg oalagaint 40week	Away Team Average goals in last 40 weeks	Result	Result of match

Table 1 Features of one instance

Since the betting companies provide odd for each outcome of a match, it makes more sense to provide probability for each outcome when doing the recommendation. After trying several different methods, developers find out that Logistic Regression can provide the probability distribution for each class label. So the Logistic Regression will be used to do the prediction analysis.

## 5.2 Data mining using Weka

Weka is used to perform the data mining task. The ten coming matches are added in front of the dataset and then the whole dataset is used as training data. In Weka, choose Logistic as classifier and choose to output prediction in more options and the prediction result is obtained (Figure 16).

=== Predictions on training set ===

```

inst#,    actual, predicted, error, probability distribution
  1         ?      2:hw      +  0.316 *0.438  0.246
  2         ?      1:draw      + *0.415  0.37  0.215
  3         ?      2:hw      +  0.336 *0.374  0.29
  4         ?      2:hw      +  0.236 *0.684  0.08
  5         ?      2:hw      +  0.277 *0.575  0.148
  6         ?      1:draw      + *0.381  0.309  0.31
  7         ?      2:hw      +  0.298 *0.546  0.156
  8         ?      2:hw      +  0.343 *0.451  0.206
  9         ?      2:hw      +  0.219 *0.692  0.089
 10        ?      2:hw      +  0.264 *0.599  0.138

```

Figure 16 Prediction result in Weka

## 5.3 Evaluation of the Prediction Model

The accuracy provided by Weka is about 85% (Figure 17).

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      2953      84.7589 %
Incorrectly Classified Instances    531       15.2411 %
Kappa statistic                    0.3153
Mean absolute error                 0.1516
Root mean squared error             0.2765
Relative absolute error              77.0265 %
Root relative squared error          88.1652 %
Total Number of Instances          3484

```

Figure 17 Accuracy in Weka

To evaluate our prediction model, developers set average odds of betting companies as benchmark. Developers calculate the average odds of different betting companies and convert the odds to probabilities of different outcomes. Then prediction results are compared to benchmark (Table 2).

Match	hwin	draw	awin	hwin	draw	awin
Napoli - Atalanta	0.65	0.21	0.13	0.546	0.298	0.156
Udinese - Bari	0.52	0.30	0.17	0.599	0.264	0.138
AS Roma - Cagliari	0.82	0.12	0.05	0.692	0.219	0.089
Inter - Chievo	0.84	0.10	0.04	0.684	0.236	0.08
Juventus - Parma	0.70	0.19	0.10	0.575	0.277	0.148
Palermo - Sampdoria	0.44	0.28	0.26	0.451	0.343	0.206
Genoa - AC Milan	0.45	0.28	0.25	0.374	0.336	0.29
Bologna - Catania	0.38	0.52	0.08	0.438	0.316	0.246
Livorno - Lazio	0.14	0.25	0.60	0.309	0.381	0.31
Fiorentina - Siena	0.68	0.21	0.10	0.37	0.415	0.215

Table 2 Comparison between prediction results and benchmark

The table shows clearly that most of the prediction results are quite similar as the average odds.

## 6. Conclusion

Betsmart provides soccer lovers a place where they can get whatever information they want. Also, it provides soccer betters a place to get guidance on how to bet and where to bet. For the future, there are several aspects that Betsmart can go ahead and improve:

1. The scale can be enlarged as it is only focused on Italian Serie A now.
2. The business model can be more versatile, e.g. develop a shopping section and bring in the online shop such as Amazon.
3. Establish the global relationship with betting companies even soccer clubs to make it a real playground for soccer enthusiasts.

## 7. Appendix

### Team Contribution

Team Member	Tasks
Ying Jin	Web Design, Data Mining, API Integration
Cai Chen	Project Manager, API Integration, Data Mining
Ximing Yu	Architecturing, Spidering, Data Pre-processing