BetStrikerz:
MIS510 Final Report

Ali Alenizi
Farah Jafar
Nikhar Shah
Yirong Zhu
# Table of Contents

INTRODUCTION 3

MARKET LANDSCAPE 3

SYSTEM OBJECTIVES 4

SYSTEM DESIGN 4

DATA SOURCES 5
DATABASES 5
ALGORITHMS 6

APIS USED 8
Facebook Login 8
Vevo 8
Twitter Feed 9
FIFA 9
FB Like, G+, Twitter Follow 10
Flickr 10
Google Map 10
World Cup Portal 11
SlideShare 11
BetStrikerz 12

NOVELTY AND UNIQUE FEATURES 12

BUSINESS MODEL 13

Promotion/ Marketing Plans: 13

REVENUE MODEL 13

Betting Companies 14
Advertisements 14
Membership 14
Fake Currency 14
Selling Features 14
Three Years Revenue Projection 14
First Year 14
Second Year 14
Third Year 14

FUTURE DIRECTIONS 15

ROLES AND CONTRIBUTIONS 15
Ali Aleinizi: 15
Farah Jafar: 16
Nikhar Shah: 16
Yirong Zhu: 16

REFERENCES 17
Introduction

The 2014 FIFA soccer world cup is just around the corner and billions of die hard soccer fans around the world are anxiously waiting for it. We have an opportunity to create a website that will bring together those soccer fans from all over the globe. Our idea is to create a website that can provide a single platform to allow fans to socialize, engage one another and most importantly bet on the upcoming World Cup 2014 matches. The betting industry has been on the rise for the past five years and is expected to grow by 18% this year (Yogonet). We have the opportunity in creating a business feasible company by taking advantage of the upcoming 2014 World Cup, large potential customer base and the fastly growing industry betting industry.

This project will serve a flow of passionate soccer fans around the globe. We will have the unique opportunity to provide a platform for soccer fans to bet on the 2014 World Cup soccer matches. We will be using a set of unique data analytics to determine likely outcomes of which team will win, qualify to further stages and potentially win the World Cup tournament. The functionality of game predictions will be based around historical match data, team and player ratings and team past world cup experiences.

Market Landscape

Betting markets around the world are worth over $417 billion and from which online channels contribute only $33.8 billion, which is only 8% of the potential market and our website will try to leverage on the this unutilized potential. Statistics suggest that there has been a 35% increase in online betting since 2003. The European market, due to less legal issues, is the most developed. Europe and UK account for more than 50% of the market share.

Our website will offer rich information not only on predicting game match results but will also share recent news and videos about the teams and players. We have three main competitors: Soccer BetUS, Topbet, BetExplorer:

**Soccer BetUS:** This website offers lots of soccer betting articles and strategies on betting based on subjective opinions. But it doesn't have enough data to prove the correctness on their assertion. Past experiences from soccer commenters may be bias and could not directly reflect results as credible as statistical data could.

**Topbet:** This website provides betting features for soccer fans, but it does not have a niche sport that it focuses on. It has some recommendations on which team to bet on. However, it doesn't have enough data analysis or historical results to prove its reliability. Additionally, this website doesn't offer many objective facts such as news, photos and videos.

**BetExplorer:** This website offers lots of betting statistics and data for soccer fans and recommending odds and stakes, but it doesn't offer betting features or utilize the gathered data to analyze future predictions.
System Objectives

With our website we look to focus on a target user. We provide a niche focus on a sport specific market. This gives our website a narrowcast focus and caters to our particular users needs and wants. Another part of the website that will help entice users to come is the simple dashboard design of the website. This design creates a feeling of personalization for the user. Our website also provide a great newly integrated platform of social betting. This is a new concept in the online betting industry. It allows for users to enjoy betting as well as provides them with all the necessary informational outlets to come to a well-informed decision for betting. We also provide a unique prediction algorithm that entails a powerful historical data analysis as well as current status integration to give final predictions.

This website will provide unique information that allow soccer fans to predict the result of 2014 World Cup Soccer matches. We would have all the historical match results between teams and recent specific data on players stored in our database. From these statistics, we would be able to calculate the overall probability of result between any two teams. The probability could be calculated by using the recursive definition based on the network of teams’ historical match results, certain scores they achieved among each other, FIFA rankings and core players’ performance. This design would apply data analysis, algorithm design, object oriented programming, and statistics method. We may use Java, Python, XML and SQL techniques in the project design.

System Design

Cloud Server:
Google APP Engine is a web server that would offer the cloud infrastructure to meet our needs. The interface supports Java and Python. We will consider using either of them.

Data Collection
We would extract data from FIFA official website, Betexplorer website, bet365 and so on. After data collection, we would use data mining skills to predict real performance among teams.

API Implementation
Latest news could be collected through ESPN, FIFA Ranking and Social Media. Graphs or photos would be available from Flickr. Videos would come from YouTube.
The above figure shows the architecture of our system. The whole system is built on Google App engine and we chose Amazon EC2 as the web framework. We collected data from FIFA website and FIFA WorldCup2014’s official website intelligent spiders were built to minimize efforts needed to collect data. The Prediction Algorithm was built on a two dimensional regression analysis with 12 years of previous data. API’s were embedded on web framework.

**Data Sources**

We apply data mining method to crawl the match results in the past 12 years on this website for any of the 2 teams within the 32 World Cup teams. We also store the matches’ information into our database.

**Databases**

After crawling the Fifa official website, we have more than 1000 historical data set in the database. The data schema has Country1,.PlayDate and Country 2 as the primary key, and Score1, Score2 and Weight as normal attributes.

Then we implement the sequel query to get the data for implementation on the prediction part:

```sql
Select avg((score1 - score2) * weight)
From countryinfo
Group by country1, country2
Order by country1
```

This is what I want to help to achieve in the algorithm part.
Algorithms

The Algorithm of this prediction model is divided into two parts. The first part is the base case, which is the current state of every team. The second part is the prediction factor that applies the two dimensional regression model based on the historical data. As the objective is to make prediction for the future state, the key point is to stand the current state and use large amount of historical data to get the trend so that the current state could be jumped to the next state. This is also similar to the principle of Markov Chain Random Process Model, but our model doesn't count too much on the time series effect.

As mentioned above, the first part should be the current state of all the teams. I apply the current Fifa ranking score for all the 32 World Cup Teams.

The second part is the prediction factor. In this part, I apply the two dimensional regression model to make the prediction:

\[
\begin{bmatrix}
2 \\
1 \\
-3 \\
2 \\
2
\end{bmatrix}, \quad \begin{bmatrix}
1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1 \\
0 & -1 & 0 & 1 \\
1 & 0 & 0 & -1 \\
0 & 1 & -1 & 0
\end{bmatrix}, \quad \begin{bmatrix}
r_A \\
r_B \\
r_C \\
r_D
\end{bmatrix}, \quad \begin{bmatrix}
e_1 \\
e_2 \\
e_3 \\
e_4 \\
e_5
\end{bmatrix}
\]

If the matrix \(X^T X\) has full rank, the algebraic solution of the system may be found via the Least squares method:

\[
r = (X^T X)^{-1} X^T y
\]

If not, one can use the Moore–Penrose pseudoinverse to get:

\[
r = X^+ y
\]

In our model, \(y\) is the matrix that stores the weighted average differences of matches in the past 12 years between any of the 32 teams. The weight is the time effect. If it's closed to 2014, the weight is higher, otherwise, the weight would be lower. We calculate the \(y\) matrix by the following procedure:

```python
def Get_Y_Matrix(c):
    Y = []
    cur = c.cursor()
    cur.execute("Select avg((score1 - score2) * weight) From countryinfo Group by country1, country2 Order by country1")
    fetch = cur.fetchall()
    for result in fetch:
        Y.append(float(result[0]))
    return Y
```

Then, the \(X\) Matrix is the coefficient factor matrix. In \(X\), in each row, it clarifies the state of the average power between the two teams. If the row value in \(Y\) gets positive, the first team in the corresponding row gets 1 and the second team gets -1 and vice versa. So, we have a 32 columns, 496 rows matrix \(X\), because the number of probable matches among the 32 teams is 496. Here is the procedure to calculate the matrix \(X\).

```python
def Get_X_Matrix(c):
```

6
Country = Get_Country_Matrix(c)
X = []
c =
c.sort()
for element in Country:
    component = []
    for i in range(0, 32):
        component.append(0)
    if element[1] == 0:
        country1 = c.index(element[0][0])
        country2 = c.index(element[0][1])
        component[country1] = 0
        component[country2] = 0
    elif element[1] > 0:
        country1 = c.index(element[0][0])
        country2 = c.index(element[0][1])
        component[country1] = 1
        component[country2] = -1
    else:
        country1 = c.index(element[0][0])
        country2 = c.index(element[0][1])
        component[country1] = -1
        component[country2] = 1
    X.append(component)
return X

After we get the matrix X and Y, we can take the value into the regression model to get the result:
def Regression_Calculation():
    Y = matrix( Get_Y_Matrix(con), dtype = float )
    X = matrix( Get_X_Matrix(con), dtype = float )
    if matrix_rank(X) == 32:
        return (X.T * X).I * X.T * Y.T
    else:
        return numpy.linalg.pinv(X) * Y.T

The above procedure gives us the final result of the ranking for the prediction factor.

Then, the last thing is to add up the historical ranking score with the current state and then we finalize the prediction model.
def Final_Rank():
    h_rank = Regression_Calculation()
    h_list = numpy.array(h_rank).tolist()
    R_rank = []
    for r in h_list:
        R_rank.append(r[0])
        R_rank.append(r[1])
for i in range(0, 32):
    Fifa_rank[i][1] += Fifa_rank[i][1] * R_rank[i]

return Fifa_rank

The final rank can help to determine that for any of the two teams, what the probability or percentage that the first team has to beat against another team.

End Product:

---

**APIs Used**

**Facebook Login**
This API uses the FB text and FB graph to output the user’s name and Facebook profile picture. We decided to use this API to give people an option of how to log into our website.

**Vevo**
This API is used to display the world cup official song on the homepage of our website. We have decided to use Vevo because it already protects against copyright infringement. The videos are also tested against crashing, ensuring that it would work on our website.
Twitter Feed
The Twitter feed API is used in the homepage to show people real time tweets made by the official FIFA World Cup account. It is their one stop twitter feed for all things FIFA!

FIFA
The FIFA APIs provide useful information for the user, and are located in the homepage. They are a set of three APIs. The first provides the latest news segments from Fifa.com. The second provides the status of different world leagues and how the different teams are doing in their respective leagues. The third provides
an opportunity to follow and track a particular world cup team and see the latest information about their rankings, matches, and news.

FB Like, G+, Twitter Follow
This is a mashup API that provides the Facebook Like API, G+ add, and the Twitter share/follow. This serves as a great way for both the users to stay connected with the website, as well as provide us with advertisement.

Flickr
flickr is used to incorporate pictures of Brazil in the WorldCup 2014 page for users to peruse as they look for information about the world cup.

Google Map
We use the Google Map API to show where we are located in the contact us page of the website. This is mainly used to allow users and investors an opportunity to know exactly where this project stemmed from.
World Cup Portal

We use the World Cup Portal API to give more information and the latest news about the World Cup and where it is hosted this year. This also provides users into the Copa2014.gov homepage without having to leave our website.

SlideShare

We use the SlideShare API in the About Us page. This is put here for potential investors to look at and discover more about where our project started and how it was developed.
The BetStrikerz API is the heart and soul of our project. This API is found on the BET page of our website. Given any two participating teams in the world cup, it will yield a probability prediction of which team will lose and which team will win.

The graph means that Brazil has 45.64% to beat against Spain, which has 54.36% to win Brazil.

This graph shows error because China is not one of the 32 World Cup teams, so it appears error to end up the prediction.

If we put same countries in two bars, we will get 50%.

Novelty and Unique Features

The main unique feature of our website is the BetStrikerz Prediction API. This is an API built by our team to predict who would win in a particular match. Thousands of data point were pulled from the FIFA website to
help build this algorithm. To achieve this algorithm we use historical data pulled from the FIFA website as well as the current state of a team, taken from the FIFA Ranking score. The historical data is integrated using a two dimensional regression analysis for 12 years of historical data.

Another unique feature of our website, are the unique APIs we use to help make our site a one stop shop for the fans and users. WE have incorporated three FIFA APIs, as well as the Copa2014.gov API and the SlideShare API.

Lastly, this entire project as a whole is a novel idea for the online betting industry. Most online gambling sites cater to all gamblers. In our website, we target a specific audience using a niche market focus. We also provide with a clear and simple business model to help get people started.

Business Model

Promotion/ Marketing Plans:

Duration: This product will focus first on the 2014 World Cup Football Match, but it will expand in the future to other venues such as the Asian Cup, European Cup or the African Cup that would only use part of our data function. We would update our database and promise the latest information for customers.

Online Guerrilla Marketing: through social media, AdSense and other relevant online venues to sports and betting.

Below is a screenshot of how this system would work:

Consumer: The consumer would purchase a service from us. This service is using the platform we have built. In return, they would be giving us monetary compensation for our services. They would also be giving us data to analyze by their transactions, as well as exposure for other consumers to use our platform.

Company: In terms of other companies, they would purchase our product (algorithm). By doing this, they increase our company's reputation and provide us with monetary compensation.

Government: Due to the nature of our company, we would need to purchase rights from the government in order for our company to be legal.

Revenue Model

Betstrikerz is a place where users can have fun as well as make money. 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 probability of a country to win in any match against any country, the betterers are very likely to choose the recommended countries. Considering all these, the betting companies and independent betterers would be happy to partner with Betstrikerz.

Advertisements
Because this website is focused on soccer games and soccer betting, the users are very likely to be soccer fans and online betterers. Therefore, it would be more effective to post soccer-related ads in this website than in others. Also we can make use of Google adsense to make the website popular.

Membership
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.

- **Different membership programs**: Gold, Silver and Bronze membership each with their own unique features.
- **Freemium model**: We would offer this product for free for some time and charge later if this product becomes popular and has a growing customers.

Fake Currency
Due to legal obligation our website cannot allow its user to bet with real currency so we can start selling the website’s own currency with which users can bet and cash out. Thus selling fake currency can generate revenue for the website. Similar concept is used by Zynga as well.

Selling Features
From customer’s perspective, the demand would be always abundant as long as we could offer accurate enough prediction. From company’s perspective, we may sell analytics, design APIs and algorithms if the function of our product is really well designed.

Three Years Revenue Projection
The revenue projection that Betstrikerz estimated breaks a profit right after the second year, which is a great outcome for a startup company.

First Year
We are assuming that we will not have a large turnout for the first year because of the World Cup 2014 rapid approach. However, we are still projecting a solid revenue of $75,000 with the 7,500 subscribers to our $10 prediction analysis charge per user. We are also hiring an outside company in the first year to develop the website frontend design and the design of the prediction API if the cost permits.

Second Year
In the second year we will be moving into the continental tournaments such as the UEFA and Asian Cup. We are assuming that the revenue changes due to increase in the market capture percentage of 5% and the increase in the subscription price to $15 as well, which is natural for price jump to happen every year. Most of the expenses incurred here are from the salaries and wages.

Third Year
In this year potentially major changes may happen. We will be adding continental soccer leagues such as the UEFA champions’ league which will add a significant amount of revenue. Additionally, in this year (2016) the Olympics Games will take place and the soccer games will have a significant impact on our revenue as well. Just like year two, we notice a market capture percentage and subscription price increase. The largest expense to incur here is the hiring of a full time developer to support the needed maintenance and website
improvements. Additionally, online marketing also increases by a significant amount from the previous years in order to ensure larger online presence and users is sustained.

Future Directions

There are a few things we look to do with the website in the future. First, we would like to enhance and monetize the prediction algorithm. By enhance we mean that as of now the GUI running our algorithm looks a little sloppy. We would need to modify it to make it more presentable. We also would want to monetize the prediction algorithm, and market it to individual users and companies. We also would want to hire a frontend developer to help optimize our frontend, and make it look more presentable. We can also do this by crowdsourcing through oDesk. Lastly, we want to ensure the future viability of our project. We can do this by moving into other target markets within the soccer scene such as the UEFA Champions League, Europa Cup, CONCACAF, Asian and European Cup Tournaments, etc.

Roles and Contributions

Ali Aleinizi:
- Backend Database Creation
- EC2 connection
- Business Model
- Flickr API
- World Cup Portal API
- SlideShare API
Farah Jafar:
- Frontend Design, Layout, Integration
- Facebook login API (FB text, FB graph)
- Twitter feed API
- FIFA API: News, FIFA World Leagues
- Market Landscape
- Competitive Analysis, and Advantage
- PPT Slides Demo 1, Demo 2
- Final Report

Nikhar Shah:
- Frontend Design, Layout
- Algorithm Data Extraction & Collection
- Data Pre-Processing
- Vevo API
- Google Map API
- PPT Slides Demo 1
- Final Report

Yirong Zhu:
- Backend Database integration
- Data analysis
- Algorithm Creation
- BetStrikerz API (creation and implementation)
- FB like, G+, Twitter Follow (Mashup) API
- FIFA API: 2014 FIFAWC team status
References

http://www.betus.com.pa/

http://www.betexplorer.com/

http://topbet.eu/

http://zynga.com/

https://developers.google.com/appengine/?csw=1

https://aws.amazon.com/ec2/