SAN JOSÉ STATE UNIVERSITY

Project Report
On

COMPANY ACQUISITION PREDICTION

CMPE 239 Web And Data Mining

By: Team 24

Vijesh Jain
Pranav Dhapke
Priyanka Bhansali
Vrushank Doshi

ACKNOWLEDGMENT

We, Team 24, take this opportunity to thank Prof. Chandrasekar Vuppalapati, for his constant support, co-ordination and encouragement. The lectures helped us in experimenting new technologies and think in a different way. We also thank him for his patience during the course of the project.

We are also grateful to our friends for their support and feedback about our Project.

ABSTRACT:

This project focuses on the future acquisition of companies in IT sector based on the companies past 3 years of financial data. In twenty first century, acquisitions have become the most frequently used methods of growth for companies. An acquisition is considered successful if the acquiring firm value is increased and both the acquiring and acquired company benefit the competition and consumers. In the world of globalization and increasing economy, top companies compete with each other to achieve optimum share in the market. Everyday people from top management level to low management level work hard to achieve the best in what they do and getting their goals as quick as possible. Some companies also follow various tactics like competing in the market of their core competencies or fighting against their rivals in the same business. But out of these, the most popular way is to take over – acquisition.

The purpose of this project is to provide a systematic analysis of the company's financial data and determine whether the company can acquire other company or can be acquired in the future. The project uses a data set which contains information of past 3 year's financial data. In this project, we have primary focused on the case study of Nokia acquisition by Microsoft in 2013.

Table of Contents

PROJECT DESCRIPTION:

One of the important factors in acquiring company is the idea of synergy that will increase the performance and decrease the costs by combining business activities. Essentially, a business will attempt to merge with another business that has complementary strengths and weaknesses. Companies often acquire other companies due to various reasons. One of them is that the acquired company may no longer have the funds needed to continue to operate effectively and decides to sell or companies may merge resources with one of them becoming parent or owner of the new company.

Companies may decide to merge in order to gain a better distribution or marketing network. A company may want to expand into different markets where a similar company is already operating rather than start from ground zero, and so the company may just merge with other company. This distribution or marketing network gives both companies a wider customer base practically overnight. The company may need to merge or get acquired by a bigger company if the current owners can't identify someone within the firm to succeed them. The owners may also wish to cash out to invest their money in something else, such as retirement. It's never easy for a company to willingly give up its identity to another company, but sometimes it is the only option in order for the company to survive.

During financial crisis, many banks merged in order to deleverage failing balance sheets that otherwise may have put them out of business. Merging of companies has some potential downslides for employees, who have to deal with immediate fears about employment or business lines, but more positive sides of merging may include more opportunities for advancement, or having access to more resources to do one's job. The great way to strategically increase revenue is by combining products and services or distribution networks. This type of acquisition or merger is highly scrutinized by Federal Trade Commission to make sure a monopoly is not created.

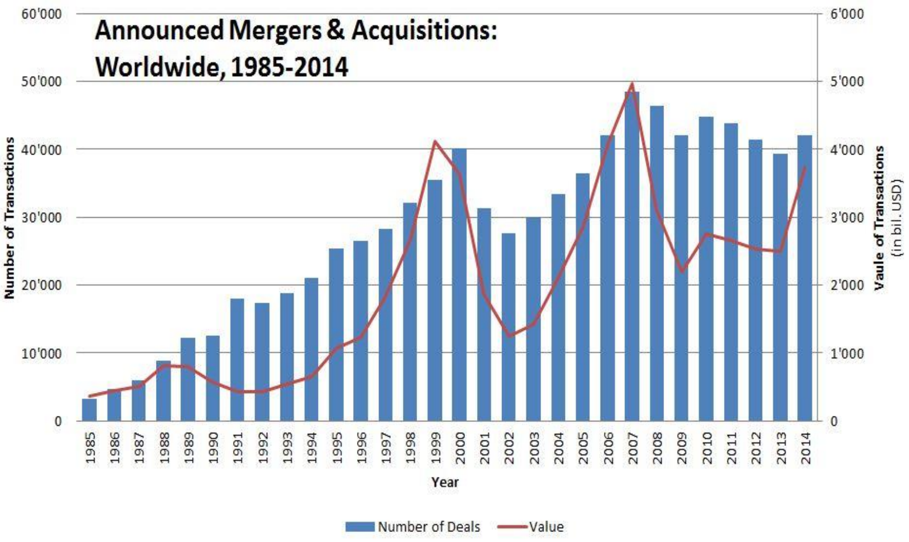


Figure 1: Mergers & Acquisitions

The term "acquisition" is used when a larger firm absorbs a smaller firm. There are many benefits behind the acquisition. Most importantly it increases value generation, increase in cost efficiency and market share. Acquisition helps in generating tax gains, can increase revenue and can reduce the cost of capital.

Industry Cases:

Sr. No.	Deal	Year	Amount in $
1	Google buys Motorola Mobility	2011	12.5 Billion
2	Microsoft buys Skype	2011	8.5 Billion
3	Microsoft-Nokia deal	2013	7.2 Billion
4	Facebook buys Instagram	2012	1.0 Billion
5	Facebook buys WhatsApp	2014	19 Billion

Table 1 – Acquisition –Merger example

Past Acquisition:

Nokia acquisition by Microsoft for $7.2 Billion in 2013.
Factors leading to this deal:

1. It will triple Windows Phone Market Share by 2018 in many markets across the globe.
2. It will drive volumes of business for Microsoft in mobile phone business domain.
3. Nokia would bring its 42000 employees on Microsoft work board which would improve its power in the mobile business.
4. To compete better with Android and iOS & also to take full control of its smartphone destiny.
5. It would position Microsoft as number three providers in the smartphone domain.

Microsoft financial statement after and before Nokia acquisition:

$ in Million	2014	2013	2012
Revenue	86,333	77,849	73,723
Operating Income	27,759	26,764	21,763
Net Income	22,074	21,863	16,978
Earning per share	2.63	2.58	2.00
Cash Dividend declared per share	1.12	0.92	0.80
Cash, Cash equivalent and short term investment	85,709	77,022	63,040
Total Asset	172,384	142,431	121,271
Long term obligation	36,975	26,070	22,220
Stockholder’s equity	89,784	78,944	66,363

Table 2 – Microsoft – Nokia financial data before and after acquisition.

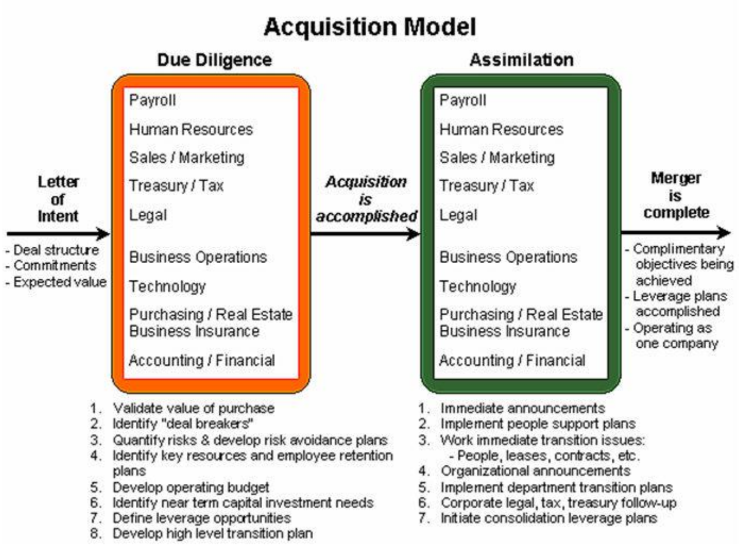

Acquisition Model
Due Diligence
Assimilation
Payroll
Human Resources
Sales / Marketing
Treasury / Tax
Legal
Business Operations
Technology
Purchasing / Real Estate
Business Insurance
Accounting / Financial
Letter of Intent
- Deal structure
- Commitments
- Expected value
Acquisition is accomplished
Merger is complete
- Complimentary objectives being achieved
- Leverage plans accomplished
- Operating as one company
1. Validate value of purchase
2. Identify "deal breakers"
3. Quantify risks & develop risk avoidance plans
4. Identify key resources and employee retention plans
5. Develop operating budget
6. Identify near term capital investment needs
7. Define leverage opportunities
8. Develop high level transition plan
1. Immediate announcements
2. Implement people support plans
3. Work immediate transition issues:
- People, leases, contracts, etc.
4. Organizational announcements
5. Implement department transition plans
6. Corporate legal, tax, treasury follow-up
7. Initiate consolidation leverage plans

Figure 2: Acquisition Model

An acquisition begins with a Letter of Intent from the acquiring company to the targeted company. When the letter of intent is accepted, due diligence begins. The acquisition model above shows each department in the company being involved in due diligence. To complete the successful merger and acquisition of two companies, it is important for each department to conduct discovery in order to develop an appropriate transition plan.

It is vitally important that the IT organization be involved in due diligence because many of the things that must take place immediately after the deal is completed will be dependent upon the IT support organization. One due diligence is completed, final negotiations take place and the companies agree to a merger.

When the deal is completed, assimilation activities begin. The organization will be busy even if the company decides to leave the acquired company alone and to operate it independently. Even so, the parent company will want to merge to Payroll, accounting and accounts payable into the parent company systems.

In addition, the company wants everyone to be on the same email systems and all company offices networked together for communication purposes, requires IT support to make it happen. The assimilation box also shows every department being involved in assimilation activity. Once the key transition projects are completed, the merger is done and they now operate as one company.

REQUIREMENTS:

1. Twitter Feeds:
   Feed pages should be present in the system to update the latest feeds continuously, based on the user preferences and search history.

2. Hashtag Sentiment Streaming:
   Streaming of tweets could be done based on keywords that are most popularly tweeted.

3. Financial Data:
   Financial Data of companies are required in order to process the data.

UI DESIGN PRINCIPLES–STORYBOARD, WIREFRAMES:

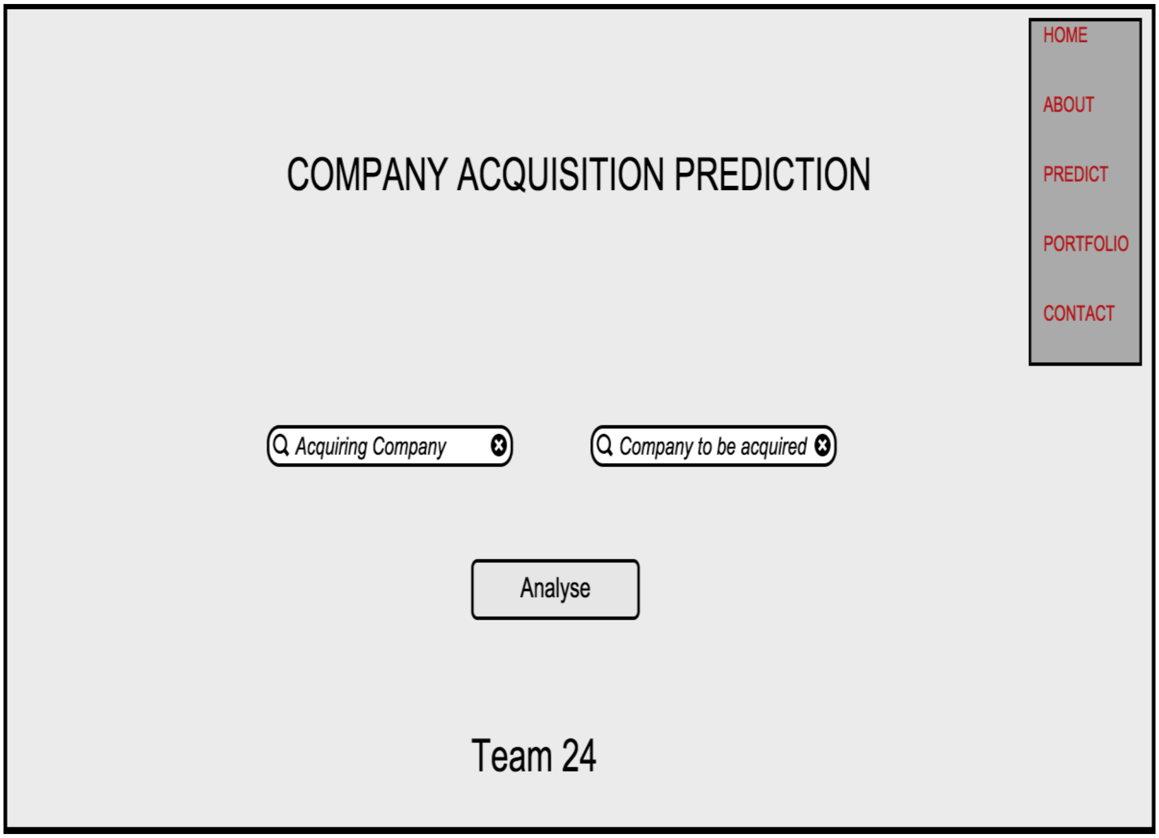


Figure 3.1 - Wireframe

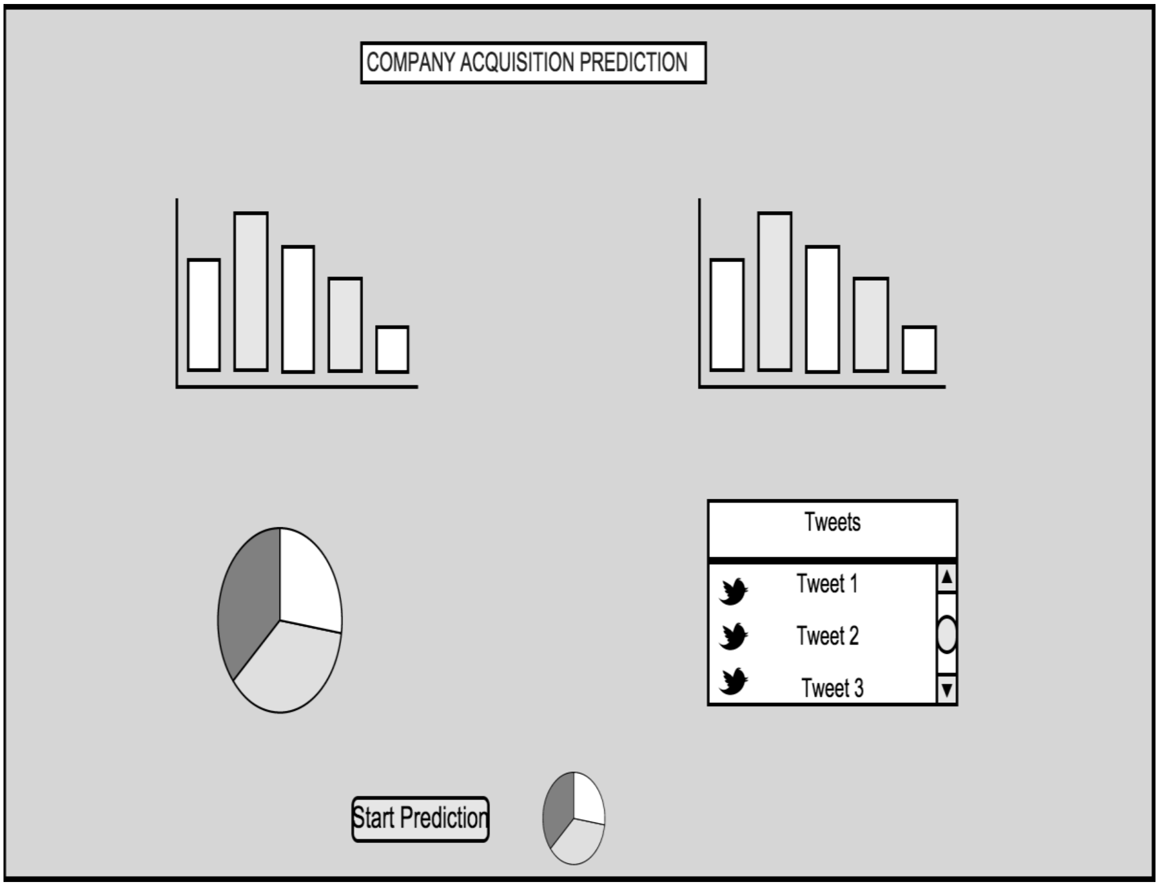


Figure 3.2 - Wireframe

Storyboard Events:

1. User browses through our website
2. Selects the Predict tab where he can enter the name of companies to be compared (Acquiring Company and Company to be Acquired)
3. Click on Analyze button.
4. HighCharts would be displayed to view the comparison stats.
5. Click on Start Prediction.
6. Bar graph is displayed with positive, negative and neutral twitter sentiments.

HIGH LEVEL ARCHITECTURE DESIGN:

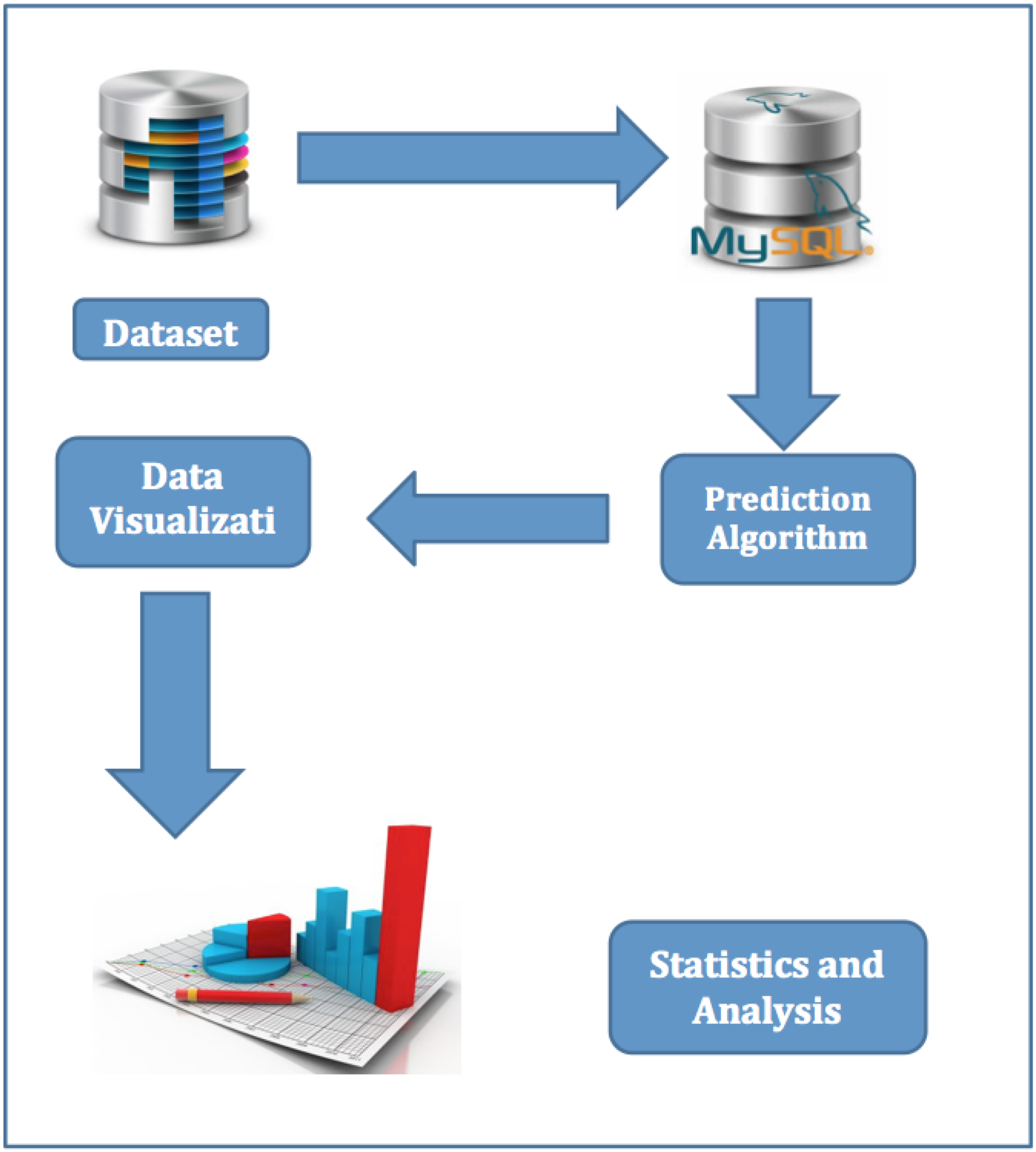
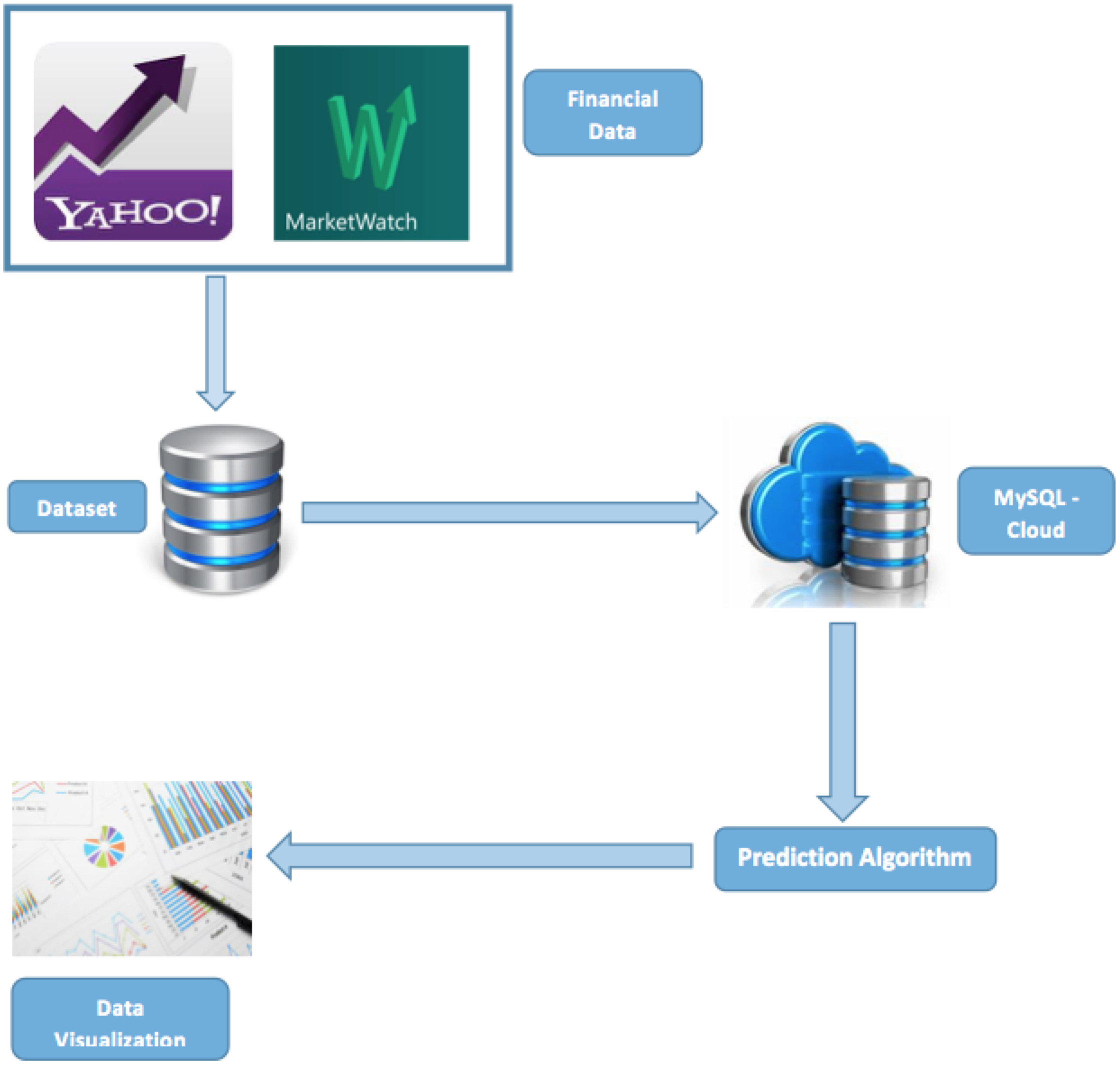


Figure 4: High Level Architecture

Description:

1. All data resides in MySQL database and is termed as dataset.
2. The data stored in MySQL is fetched to perform computation.
3. The predictive algorithm runs on the fetched data.
4. The algorithm scans through all the fetched data and predicts the result.
5. The result is passed to the visualization tool.
6. Using high charts, the data is represented on our web application.

DATASETS AND DATA PATTERNS:



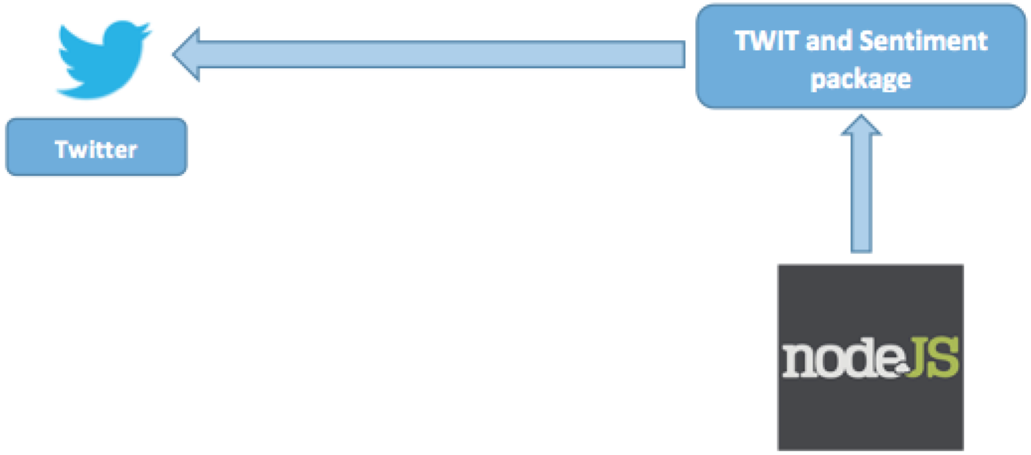


Figure 5: Data Patterns

DATA ARCHITECTURE:

Pre-processing

Tweets

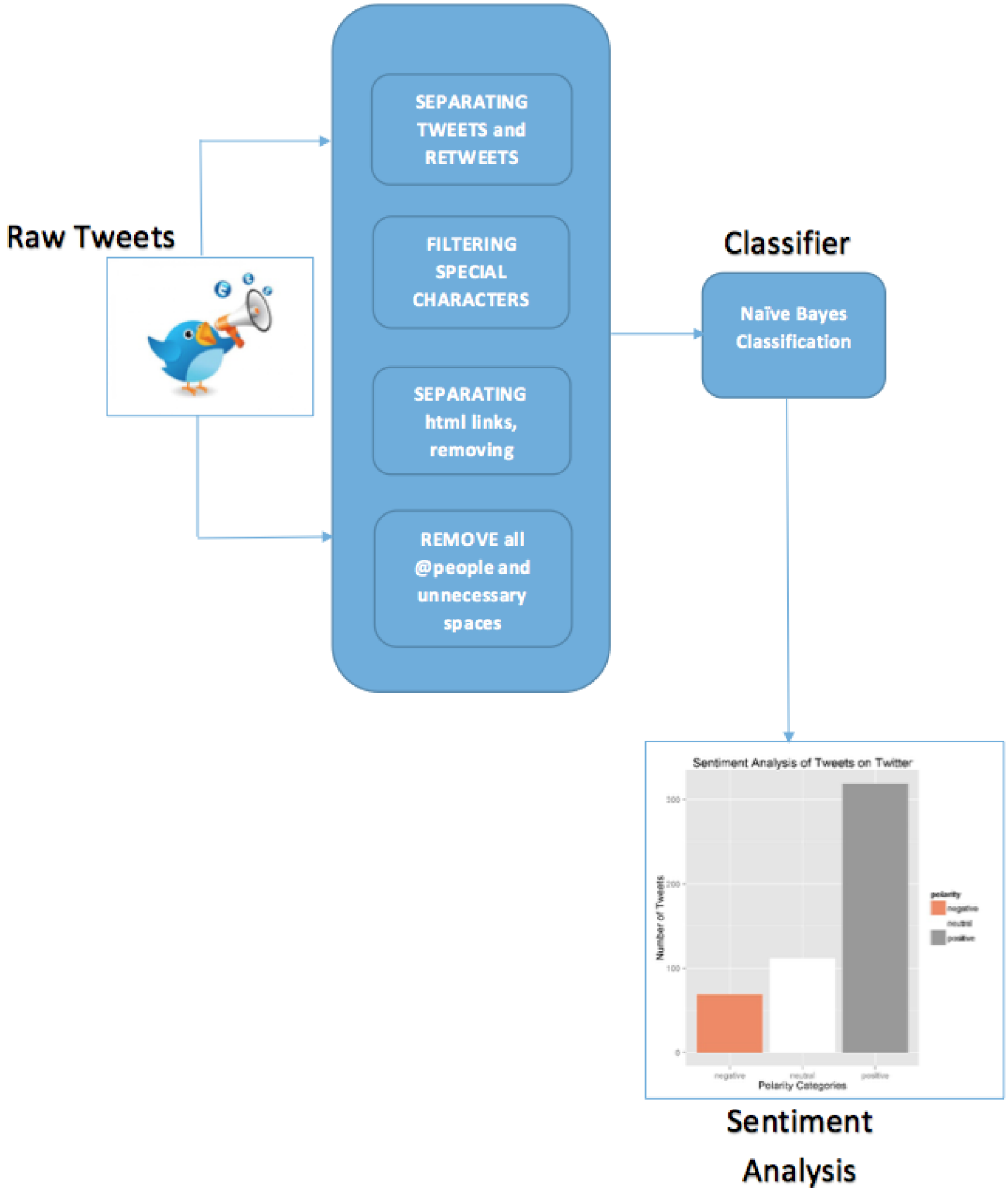

Raw Tweets
SEPARATING TWEETS and RETWEETS
FILTERING SPECIAL CHARACTERS
SEPARATING html links, removing
REMOVE all @people and unnecessary spaces
Classifier
Naïve Bayes Classification
Sentiment Analysis of Tweets on Twitter
Number of Tweets
300
200
100
0
negative
neutral
positive
Polarity Categories
polarity
negative
neutral
positive
Sentiment Analysis

Figure 6: Architecture for twitter sentiment analysis in R.

Description:

1. Perform twitter authentication to access twitter data.
2. Fetch tweets from twitter based on the user input.
3. Filter out the tweets since we only require text.
4. Remove all retweets from the result.
5. Remove all special characters and digits.

6. Remove http links and unnecessary spaces from the result.
7. Remove all tagging i.e @people.
8. This will give us a pre-processed data i.e. plain text.
9. Apply Naïve Bayes algorithm on the result.

This algorithm displays a bar graph with positive, negative and neutral sentiments as shown above.

DATA FLOW DIAGRAM:

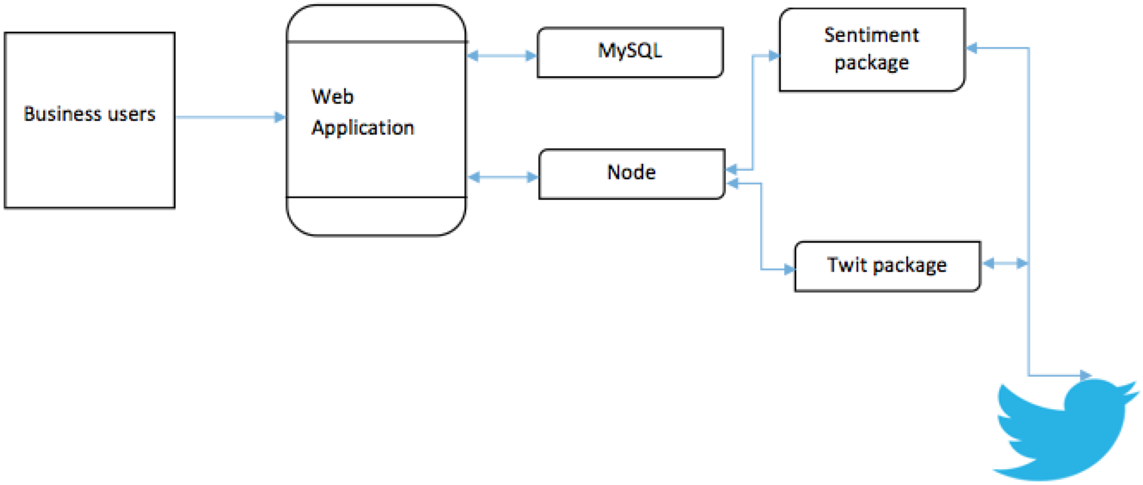


Figure 7: Data flow Diagram

Twitter Sentiment analysis using R.

```
library(ggplot2)
library(ROAuth)
library(Rstem)
library(sentiment)
library(tm)
library(twitteR)
library(plyr)
library(RColorBrewer)
library(wordcloud)


reqURL <- "https://api.twitter.com/oauth/request_token"
accessURL <- "http://api.twitter.com/oauth/access_token"
authURL <- "http://api.twitter.com/oauth/authorize"
consumerKey <- "WhYExrBdjaQKyArMvJhxjHiHt"
consumerSecret <-
"WZ81G6ADuk2RJgGRVMsRZMGgJcUudbC7W2xdhTTx1Miqbhr68y"
ACCESS_TOKEN <- "2433226039-
ALNzRIdqRR3StA2GZSXyPDuMJMhp5rm604BWktz"
ACCESS_secret <-
"nwUXQJsFraf4wd2SB8ls3BKQiB5uAWVEWO1Ea6cLk9ifB"

#twitCred <-
OAuthFactory$new(consumerKey=consumerKey,consumerSecret=consume
rSecret,requestURL=reqURL,accessURL=accessURL,authURL=authURL)
#download.file(url="http://curl.haxx.se/ca/cacert.pem",
destfile="cacert.pem")
#twitCred$handshake(cainfo="cacert.pem")
setup_twitter_oauth(consumerKey,consumerSecret, ACCESS_TOKEN,
ACCESS_secret)

args <- commandArgs(trailingOnly = TRUE)

company_tweets = searchTwitter("fb", n=500, lang="en")
company_txt = sapply(company_tweets, function(x) x$getText())
company_txt = gsub("(RT|via)((?:\\b\\W*@\\w+)+)", "", company_txt)
company_txt = gsub("@\\w+", "", company_txt)
company_txt = gsub("[[:punct:]]", "", company_txt)
company_txt = gsub("[[:digit:]]", "", company_txt)
company_txt = gsub("http\\w+", "", company_txt)
company_txt = gsub("[ \t]{2,}", "", company_txt)
company_txt = gsub("^\\s+|\\s+$", "", company_txt)
```

```
catch.error = function(x)
{
y = NA
catch_error = tryCatch(tolower(x), error=function(e) e)
if (!inherits(catch_error, "error"))
y = tolower(x)
return(y)
}

company_txt = sapply(company_txt, catch.error)
company_txt = company_txt[!is.na(company_txt)]
names(company_txt) = NULL
company_class_emo = classify_emotion(company_txt, algorithm="bayes",
prior=1.0)
emotion = company_class_emo[,7]
emotion[is.na(emotion)] = "unknown"

company_class_pol = classify_polarity(company_txt, algorithm="bayes")
polarity = company_class_pol[,4]

sentiment_dataframe = data.frame(text=company_txt, emotion=emotion,
polarity=polarity, stringsAsFactors=FALSE)
sentiment_dataframe = within(sentiment_dataframe, emotion <-
factor(emotion, levels=names(sort(table(emotion), decreasing=TRUE))))


#ggplot(sentiment_dataframe, aes(x=emotion)) + geom_bar(aes(y=..count..,
fill=emotion)) +
#scale_fill_brewer(palette="Dark2") +
#ggtitle('Sentiment Analysis of Tweets on Twitter') +
#theme(legend.position='right') + ylab('Number of Tweets') + xlab('Emotion
Categories')

ggplot(sentiment_dataframe, aes(x=polarity)) +
geom_bar(aes(y=..count.., fill=polarity)) +
scale_fill_brewer(palette="RdGy") +
ggtitle('Sentiment Analysis of Tweets on Twitter') +
theme(legend.position='right') + ylab('Number of Tweets') + xlab('Polarity
Categories')
```

```
ggsave("F:\\SentimentAnalysis1.jpg")
dev.off()

mach_corpus = Corpus(VectorSource(company_txt))
tdm = TermDocumentMatrix(mach_corpus, control =
list(removePunctuation = TRUE,stopwords = c("machine", "learning",
stopwords("english")),removeNumbers = TRUE, tolower = TRUE))
m = as.matrix(tdm)
word_freqs = sort(rowSums(m), decreasing=TRUE)
dm = data.frame(word=names(word_freqs), freq=word_freqs)

wordcloud(dm$word, dm$freq, random.order=FALSE, colors=brewer.pal(8,
"Dark2"))
jpeg("F:\\MachineLearningCloud.jpg", width=12, height=8, units="in",
res=300)
wordcloud(dm$word, dm$freq, random.order=FALSE, colors=brewer.pal(8,
"Dark2"))
dev.off()
```

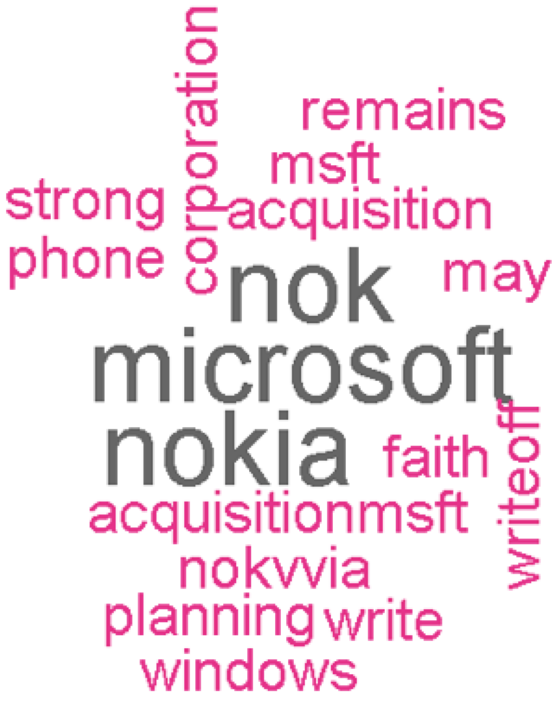


Fig. WORD CLOUD – MSFT NOKIA Acquisition

Figure 8: Word Cloud

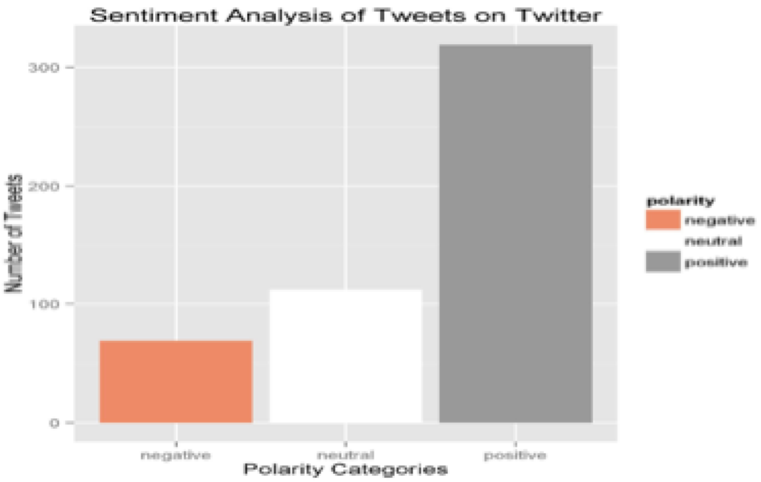


Fig. Sentiment Analysis on MSFT NOKIA Acquisition

Figure 9: Sentiment Analysis on MSFT NOKIA Acquisition

DATA MINING PRINCIPLES AND ALGORITHMS:

Company Acquisition Prediction- As the name of our project indicates, the Data Model used for our project is Predictive Model.

What do we mean by Predictive Analytics?

Predictive analytics is not the actual outcome that would take place in future. It is just a method or form of extracting/generating useful information, from the existing data sets, so as to regulate patterns that could help us determine, or in other words, 'predict' the future trends and outcomes.

For this purpose of analyzing and predicting the unknown future trends, we make use of variety of techniques from data mining, modeling and machine learning. These techniques understand and analyze both, the past and current data, to make analysis and predictions for future.

Ample of industries make use of this predicated analysis in there day-to-day work and its demand is increasing in other fields as well.

Marketing, health-cares, travel, retailers, insurance, fraud detection and many other fields have also realized the usefulness of predictive analysis.

Predictive Modeling or Predictive analytics is nothing but an area of data mining where its objective is to extract useful knowledge from the available data(past and present) and use the same knowledge for identifying patterns and predict behavior of future.

Why Use Predictive Models?

As mentioned earlier, Predictive Models are used to predict the future outcome by performing comparison of the values with huge data sets. In our project, comparison is made between two companies based on their gross profit and share values, number of employees and other elements.

From the outcome, we are able to predict if there are any possibilities of the company to be acquired.

Benefits of Predictive Modeling:

It enhances the efficiency and truthfulness of our predictions

Helps us get accurate and reliable results.

Helps in building efficient models that work more with the limited data provided

Reduces pain of adjusting and calculating values each time, saving time.

It provides better diagnostics that help in better managing, controlling and decision-making process.

Naive Bayesian Classifier:

The Naive Bayes algorithm is a classification algorithm based on Bayes' theorems, and used in predictive modeling. Naive Bayes is a formula which represents conditional probabilities of occurrence of an event A, given the occurrence of an event B. It is represented as P(A|B). The theorem states that in order to determine the probability of this condition, we need the probability of the occurrence of exactly opposite event and all individual probability of the occurrence of two elements.

i.e. $P(A|B) = P(A)P(B|A) / P(B)$

This algorithm is less computationally intense and therefore is useful for quickly generating mining models to discover relationships between input columns and predictable columns. We can use this algorithm to do initial exploration of data, and then later apply the results to create additional mining models with other algorithms that are more computationally intense and more accurate. In our case, if we wish to identify if a given tweet is positive or negative, given its contents, we can use Bayes theorem to state that the probability of occurrence of a given tweet provided that its predetermined to be either positive or negative. This is very convenient for the purpose of our calculation as we already have available examples of positive and negative tweets based upon our existing data set of tweets. This means that we are making a very broad assumption that the probability of the occurrence.

Acquisition Prediction Algorithm:

We have used following terms in making the prediction for acquisition:

Category	Equity	Total Assets	Total Liability
Operating Margin Ratio	Current Asset	Gross Profit	Share Value
Debt Equity	EBITDA	Total Revenue	
Full time Employee	Gross Margin	Cash and Cash Equivalence	

Algorithm:

1. Fetching values for the above terms from the dataset for specified company.
2. Performing calculation based on the following formula:

   Equity = Total Asset – Total Liability
   Debt Equity = Total Liability / Equity
   Operating Margin Ratio = EBITDA / Total Revenue.
3. Perform comparison of these values between the company who is acquiring and the company to be acquired.
4. Based on the result of each comparison, calculate percentage of the current year.
5. Repeat the above steps for last 3 years values.
6. Calculate average of these percentage = prediction %.

KDD PRINCIPLES:

KDD or Knowledge Discovery in Databases is a multidisciplinary branch of science that deals with data storage and data access, algorithms that are highly scalable and has huge data sets. The processes that are usually included in KDD are: Selection, Processing, Transformation, Data Mining and Evaluation. The overall process of finding and interpreting patterns from data involves recursive application of these steps.

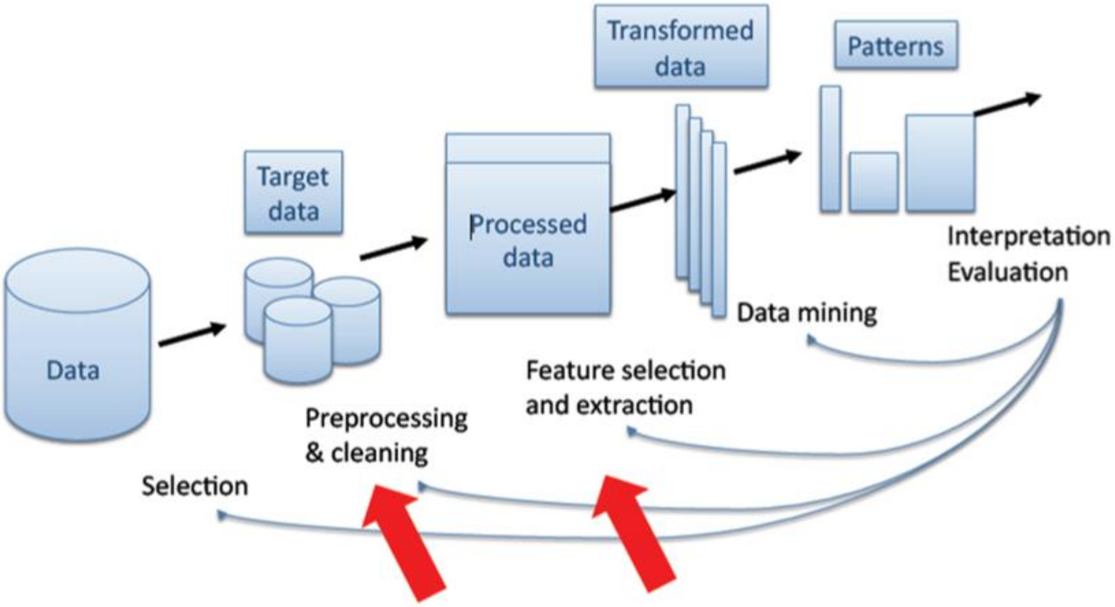


Figure 10: KDD Model

Here is the brief description of the steps involved in the KDD process:

7. Customer's goal and objectives of the KDD process has to be identified and known.
8. The idea of having knowledge of the application domain is required.
9. After selecting a target data set, it is required to create a data set on

which we would be performing the discovery.

10. Perform pre-processing of data.
11. Perform cleansing operation in order to filter out data by using appropriate patterns and strategies. Remove the variable that are not of much interest and analyze the data.
12. Compare the methods available for Data Mining techniques with the final goal to find out the matching pattern that could be merged from the KDD process.
13. Research about the algorithms that can be used for data mining in order to find out the hidden patterns.
14. Also find out the relevant patterns which includes rules and testing of the processes.
15. Interpret important data and other information required from the mined patterns.

This knowledge can be used for further processing of data.
Make a note of these observations and conclusions to prepare a document that can be put forward to different stakeholders.

We have been using KDD approach in our project for extracting useful information (tweet feeds) from the database. This knowledge involved exploring and analyzing large amounts of data to find patterns for big data based on positive, negative, neutral sentiments. Each tweet belongs to either of these sentiment classes. Inclusion of the neutral class has been in the model in order to classify those tweets that did not belong to positive or negative class.

All the steps used and applied were in accordance with KDD principles. These steps were observed, analyzed and put to use to arrive at a conclusion of getting clean data. Here are the steps that were evenly followed in our project on similar grounds after which we were able to classify the tweet feeds (positive, negative, neutral)

- Separating Tweets and Retweets (Pre-processing),
- Filtering special characters (Cleaning),
- Naïve Bayes Classification for Sentiment analysis of tweets on Twitter (Feature Selection and Extraction),
- Separating html links, removing hashtags # from the data used(Data Mining),
- Remove all @people and unnecessary spaces (Interpretation and

Evaluation)

DATA TOOLS:

Data Analytics using R:

R is a very good tool which translates the derived model into colorful graphs and visualizations. R has a drawback that it is not very scalable.
R provides wide variety of machine learning (clustering, classification, linear and non-linear modelling) and graphical techniques. R has various built-in and extended functions for statistical and machine learning tasks such as

a) Data extraction
b) Data cleaning
c) Data loading
d) Data transformation
e) Predictive modelling
f) Data visualization

With its growing list of packages, R can now connect with other data stores such as MySQL, SQLite, Hadoop and MongoDB for data storage activities. R allows performing Data Analytics by various statistical and machine learning operation as follows:-

a) Regression
b) Classification
c) Clustering
d) Recommendation
e) Text Mining.

High Charts

High Charts are used to create interactive charts on the web. We have used high charts for data visualization. Highcharts is a charting library in JavaScript, offering us an easy way of adding interactive charts to our web site or web application. Highcharts supports line, spline, area, areaspline, column, bar, pie, scatter, angular gauges, arearange, areasplinerange, columnrange, bubble, box plot, error bars, funnel, waterfall and polar chart types. We have used high charts for plotting bar graphs and pie charts.

DESIGN PATTERNS

Node JS Express MVC framework:

Express is a minimal and flexible Node.js web application framework that provides a robust set of features for web application. We have Node JS express MVC framework since it provides a lot of in built packages for different functionalities which we plan to achieve. Since it has a great support and bunch of features, Express is one of the best framework for Node.

Front End – EJS:

EJS stands for Embedded JavaScript, it is essentially HTML with JavaScript based features embedded for templating. EJS is primarily HTML but with additional features which allows us to reuse pieces of our components. If we have any HTML project, then all we need to do is to rename it to .ejs extension and we can start using EJS.

For Highchart-1

```
        $('#container1').highcharts({
                chart: {
                   backgroundColor: '#EDEEFA',
                   type: 'column'
                },
                title: {
                   text:          'Stock          Symbol:          <%=
comp1[0].Company_Name.toString().toUpperCase() %>'
                },
                /*  subtitle: {
                 text: 'Source: WorldClimate.com'
                 }, */

                xAxis: {
                   categories: [
                      'Current Assests',
                    'Liability',
                    'Gross Profit',
                    'Total Revenue',
                    'EBITDA',
                    'Gross Margin',
                    'Cash Equivalent'

                ],
                crosshair: true
            }
            ,
            yAxis: {
                min: 0,
                title: {
                    text: 'Amount (billions)'
                }
            }
            ,
            tooltip: {
                headerFormat: '<span style="font-size:10px">{point.key}</span><table>',
                pointFormat: '<tr><td style="color:{series.color};padding:0">{series.name}: </td>' +
                '<td style="padding:0"><b>{point.y:.1f} bn</b></td></tr>',
                footerFormat: '</table>',
                shared: true,
                useHTML: true
            }
            ,
            plotOptions: {
                column: {
                    pointPadding: 0,
                    borderWidth: 0
                }
            }
            ,
            series: [{
                name: "<%= comp1[0].Year %>",
                data: [
                    <%= comp1[0].Total_Current_Assets %>,
                    <%= comp1[0].Total_Liabilities %>,
                    <%= comp1[0].Gross_Margin %>,
                    <%= comp1[0].total_revenue %>,
                    <%= comp1[0].EBITDA %>,
                    <%= comp1[0].Cash_and_Cash_Equivalents %>
                ]

            }, {

                name: "<%= comp1[1].Year %>",
                data: [
                    <%= comp1[1].Total_Current_Assets %>,
                    <%= comp1[1].Total_Liabilities %>,
                    <%= comp1[1].Gross_Margin %>,
                    <%= comp1[1].total_revenue %>,
                    <%= comp1[1].EBITDA %>,
                    <%= comp1[1].Cash_and_Cash_Equivalents %>
                ]
            },
                {

                    name: "<%= comp1[2].Year %>",
                    data: [
                        <%= comp1[2].Total_Current_Assets %>,
                        <%= comp1[2].Total_Liabilities %>,
                        <%= comp1[2].Gross_Margin %>,
                        <%= comp1[2].total_revenue %>,
                        <%= comp1[2].EBITDA %>,
                        <%= comp1[2].Cash_and_Cash_Equivalents %>
                    ]
                }]
        }
);
```

Middle tier – Express:

We have used Express as a middleware for Node JS framework. An express application is a series of middleware calls. Middleware is a function with access to the request object(req), response object(res) and the next middleware in line in the request-response cycle of an Express application. Middleware can do the following:

1. Execute code
2. Make changes to the request
3. End the request response cycle
4. Call the next middleware in the stack

An express application can use following kinds of middleware:

1. Application level middleware
2. Router level middleware
3. Error handling middleware
4. Built in middleware
5. Third party middleware

We have used application level middleware. Application level middleware are bound to an instance of express.

App.js

```
var express = require('express');
var path = require('path');
var logger = require('morgan');
var cookieParser = require('cookie-parser');
var bodyParser = require('body-parser');

var routesBase = require('./routes/index');
var analysis = require('./routes/analysis');
var app = express();

// view engine setup
app.set('views', path.join(__dirname, 'views'));
app.set('view engine', 'ejs');

app.use(logger('dev'));
app.use(bodyParser.json());
app.use(bodyParser.urlencoded({extended: false}));
app.use(cookieParser());
app.use(express.static(path.join(__dirname, 'public')));

app.use('/', routesBase);

app.post('/getStats', analysis.getStats);

// catch 404 and forward to error handler
app.use(function (req, res, next) {
   var err = new Error('Not Found');
   err.status = 404;
   next(err);
});

// error handlers
// development error handler
// will print stacktrace
if (app.get('env') === 'development') {
    app.use(function (err, req, res, next) {
        res.status(err.status || 500);
        res.render('error', {
            message: err.message,
            error: err
        });
    });
}

// production error handler
// no stacktraces leaked to user
app.use(function (err, req, res, next) {
    res.status(err.status || 500);
    res.render('error', {
        message: err.message,
        error: {}
    });
});

module.exports = app;
```

Analysis.js

```
var mysql = require('./mysql');
var tweetStats = require('./twitterSearch');

exports.getStats = function (req, res) {

    var compFirst = req.param("comp1").toLowerCase().trim();
    var compSecond = req.param("comp2").toLowerCase().trim();
   var getStats = "select * from mytable1 where Company_Name='" +
compFirst + "' OR Company_Name='" + compSecond + "'";
    var polarity = [];
    tweetStats.getTweets(compFirst, compSecond, function (err, result, twits)
{
        if (err) {

            throw err;
        }

        polarity = result;
        var tweets = twits;
        console.log("hello pol..." + polarity);

        mysql.fetchData(getStats, function (err, rows) {
            if (rows.length == 0) {

                res.render('index', {error: "Error"});
                console.log(err);

            }
            else {
                console.log("pol..." + polarity);
                var comp1 = [];
                var comp2 = [];
                for (var i = 0; i < rows.length; i++) {
                  if (rows[i].Company_Name === compFirst) {
                    comp1.push(rows[i]);
                  }
                  else if (rows[i].Company_Name === compSecond) {
                    comp2.push(rows[i]);
                  }

                }

                var EV1 = [];
                var EV2 = [];
                //var asset_ratio1 = [];
                //var asset_ratio2 = [];
                var operating_MarginRatio1 = [];
                var operating_MarginRatio2 = [];
                var debt_equity1 = [];
                var debt_equity2 = [];
                var year = [];


                for (var i = 0; i < 3; i++) {

                  var counter = 0;
                  if (comp1.sector == comp2.sector) {
                    console.log("in sector");
                    counter = counter + 1;
                    console.log(counter);
                  }

                  console.log("at Total_Assets :" + comp2[i]);
                  console.log("at Total_Liabilities :" +
comp2[i].Total_Liabilities);

                  EV1.push((comp1[i].Total_Assets -
comp1[i].Total_Liabilities));
```

```
                EV2.push((comp2[i].Total_Assets                                  -
comp2[i].Total_Liabilities));
                console.log("at EV 2 :" + EV2[i]);

                debt_equity1.push((comp1[i].Total_Liabilities / EV1[i]));
                debt_equity2.push((comp2[i].Total_Liabilities / EV2[i]));
                //asset_ratio1.push((comp1[i].Total_Assets                          /
comp1[i].Total_Current_Assets));
                //asset_ratio2.push((comp2[i].Total_Assets                          /
comp2[i].Total_Current_Assets));
                operating_MarginRatio1.push((comp1[i].EBITDA                    /
comp1[i].total_revenue));
                operating_MarginRatio2.push((comp2[i].EBITDA                    /
comp2[i].total_revenue));

                console.log("value        Total_Current_Assets        1.."        +
comp1[i].Total_Current_Assets);
                console.log("value        Total_Current_Assets        2.."        +
comp2[i].Total_Current_Assets);
                console.log("Company 1: " + JSON.stringify(comp1));
                console.log("Company 2: " + JSON.stringify(comp1))

                if (EV1[i] > 2 * EV2[i]) {
                    counter = counter + 1;
                    console.log("in EV.." + counter);
                }
                if (debt_equity1[i] > 2 * debt_equity2[i]) {
                    counter = counter + 1;
                    console.log("in debt_equity.." + counter);

                }
                //if (asset_ratio1[i] > 3 * asset_ratio2[i]) {
                //    counter = counter + 1;
                //    console.log("in asset_ratio.." + counter);
                //}
                if (operating_MarginRatio1[i] > operating_MarginRatio2[i]) {
                    counter = counter + 1;
                    console.log("in operating_MarginRatio1.." + counter);
```

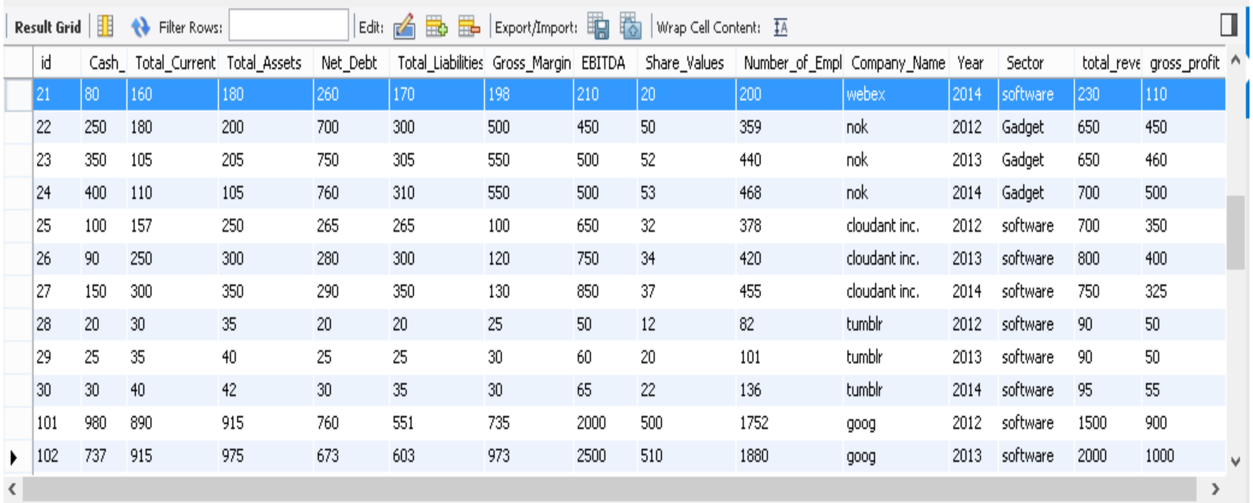
```
          }
          if (comp1[i].Total_Current_Assets > (3 *
comp2[i].Total_Current_Assets)) {
            console.log("in asset");
            counter = counter + 1;
            console.log(counter);
          }
          if (comp1[i].Total_Liabilities > (2 *
comp2[i].Total_Liabilities)) {
            console.log("in gross");
            console.log("in total");
            counter = counter + 1;
            console.log(counter);
          }
          if (comp1[i].gross_profit > 3 * comp2[i].gross_profit) {
            counter = counter + 1;
            console.log(counter);
          }
          if (comp1[i].Share_Values > comp2[i].Share_Values) {
            console.log("in share");
            counter = counter + 1;
            console.log(counter);
          }
          if (comp1[i].Number_of_Employees >
comp2[i].Number_of_Employees) {
            console.log("in full");
            counter = counter + 1;
            console.log(counter);
          }
          if (comp1[i].Gross_Margin > 3 * comp2[i].Gross_Margin) {
            console.log("in gross margin");
            counter = counter + 1;
            console.log(counter);
          }
          if (comp1[i].Cash_and_Cash_Equivalents > 2 *
comp2[i].Cash_and_Cash_Equivalents) {
            console.log("in cash");
            counter = counter + 1;
```

```
                        console.log(counter);
                    }
                    year.push(Math.round((counter / 11) * 100));
                }
                console.log("y1: " + year[0] + " y2: " + year[1] + " y3: " +
year[2]);
                var finalPercent = parseInt(((year[0] + year[1] + year[2]) / 3));
                console.log("final percent: " + finalPercent);

                //console.log(JSON.stringify(rows));
                res.render('viewStats', {
                    data: rows,
                    comp1: comp1,
                    comp2: comp2,
                    polarity: polarity,
                    tweets: tweets,
                    finalPercent: finalPercent
                });
            }
            // render or error
        });
    });
}
```

Data Store:

MySQL is used for data storage. We have created a mock data for a set of companies. It is a relational database that stores data in the form of rows and columns i.e. tabular representation of data.

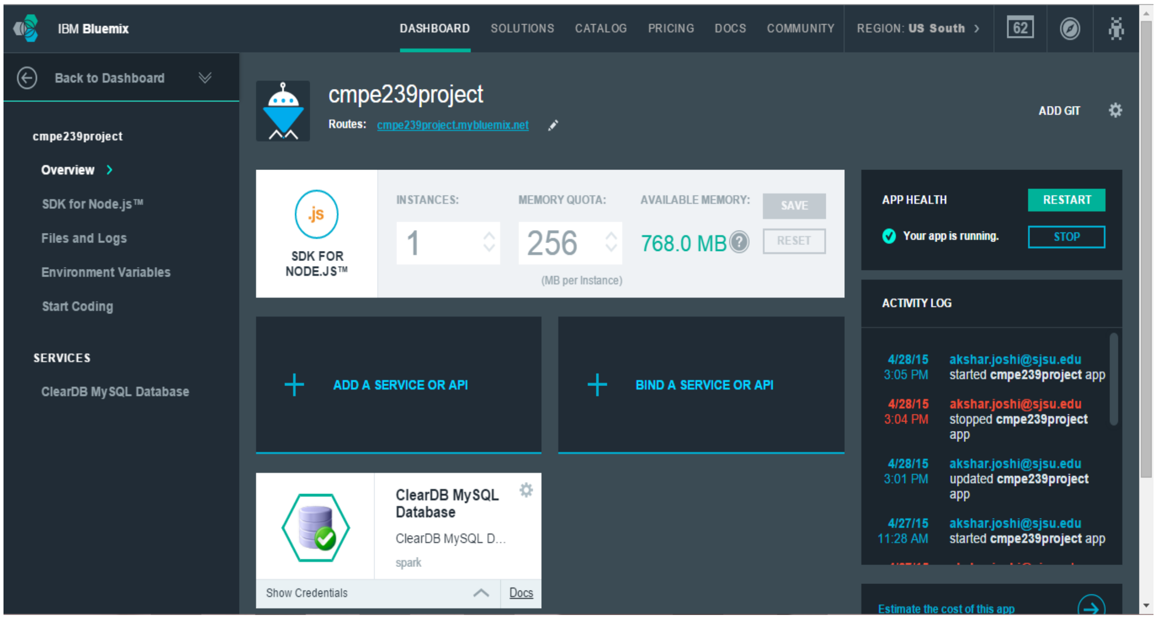


id	Cash_	Total_Current	Total_Assets	Net_Debt	Total_Liabilities	Gross_Margin	EBITDA	Share_Values	Number_of_Empl	Company_Name	Year	Sector	total_reve	gross_profit
21	80	160	180	260	170	198	210	20	200	webex	2014	software	230	110
22	250	180	200	700	300	500	450	50	359	nok	2012	Gadget	650	450
23	350	105	205	750	305	550	500	52	440	nok	2013	Gadget	650	460
24	400	110	105	760	310	550	500	53	468	nok	2014	Gadget	700	500
25	100	157	250	265	265	100	650	32	378	cloudant inc.	2012	software	700	350
26	90	250	300	280	300	120	750	34	420	cloudant inc.	2013	software	800	400
27	150	300	350	290	350	130	850	37	455	cloudant inc.	2014	software	750	325
28	20	30	35	20	20	25	50	12	82	tumblr	2012	software	90	50
29	25	35	40	25	25	30	60	20	101	tumblr	2013	software	90	50
30	30	40	42	30	35	30	65	22	136	tumblr	2014	software	95	55
101	980	890	915	760	551	735	2000	500	1752	goog	2012	software	1500	900
102	737	915	975	673	603	973	2500	510	1880	goog	2013	software	2000	1000

Figure 11: Data store in MySQL

Cloud:

We have used IBM Bluemix for hosting our web application and database. It is a open standards, cloud platform for building, running and managing apps and services. IBM Bluemix is a cloud platform that helps developers rapidly build, manage and run web as well as mobile applications.



cmpe239project

Routes: cmpe239project.mybluemix.net

INSTANCES: 1

MEMORY QUOTA: 256 (MB per instance)

AVAILABLE MEMORY: 768.0 MB

Your app is running.

ClearDB MySQL Database
ClearDB MySQL D...
spark

4/28/15 3:05 PM akshar.joshi@sjsu.edu started cmpe239project app

4/28/15 3:04 PM akshar.joshi@sjsu.edu stopped cmpe239project app

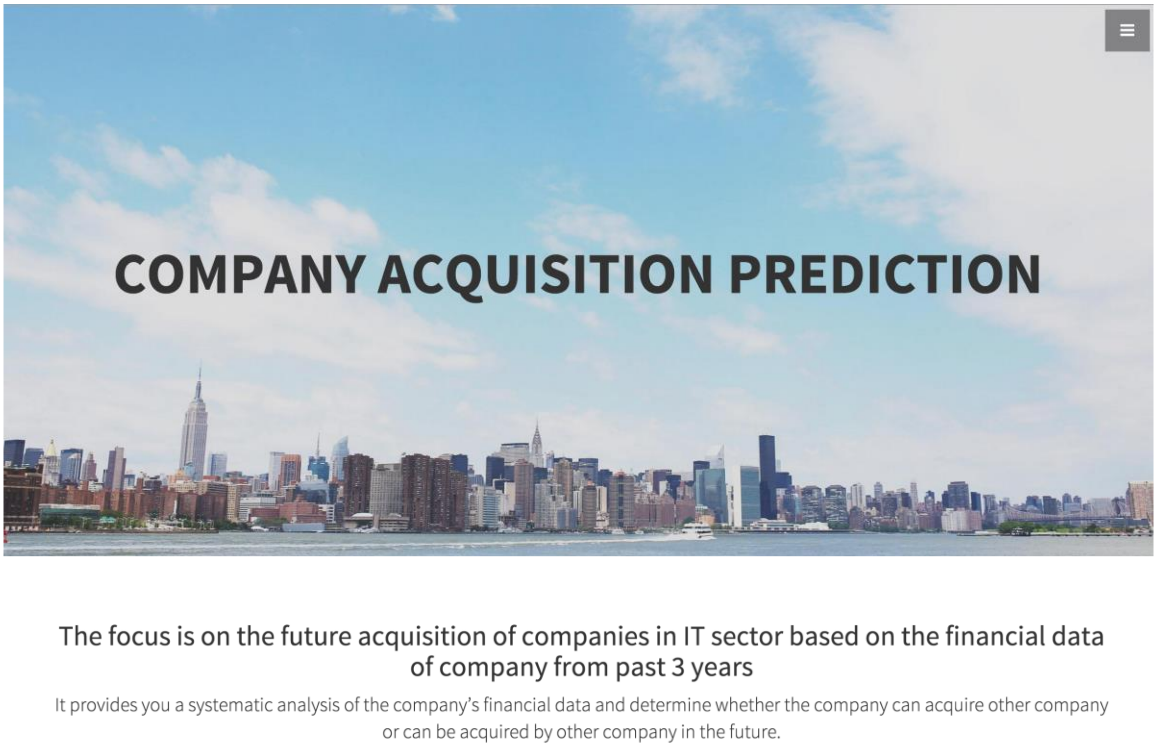
4/28/15 3:01 PM akshar.joshi@sjsu.edu updated cmpe239project app

4/27/15 11:28 AM akshar.joshi@sjsu.edu started cmpe239project app

Figure 12: Screenshot from IBM Bluemix

CLIENT SIDE DESIGN:

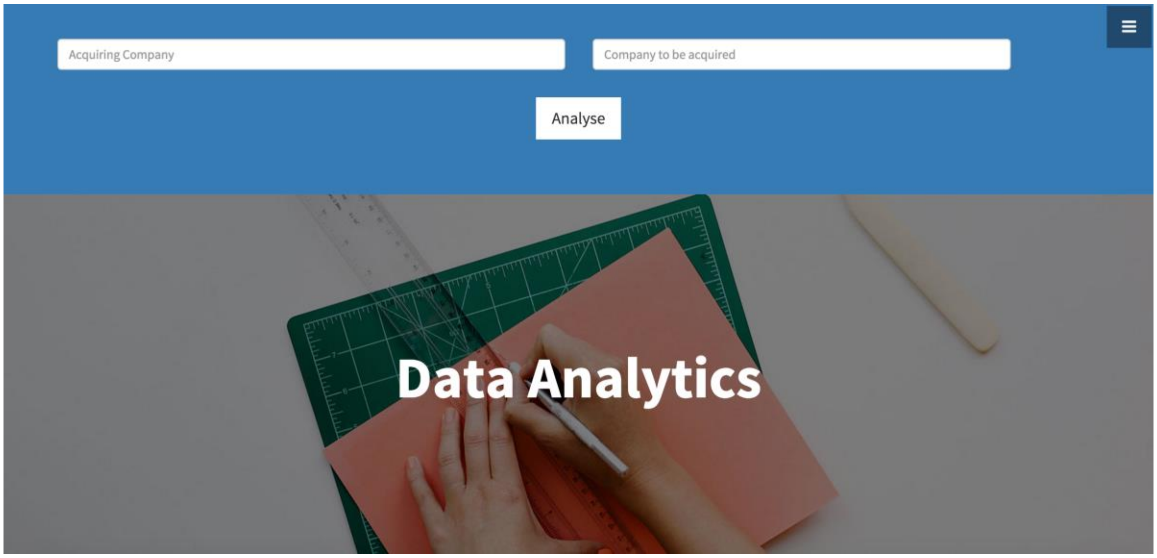
Home Page

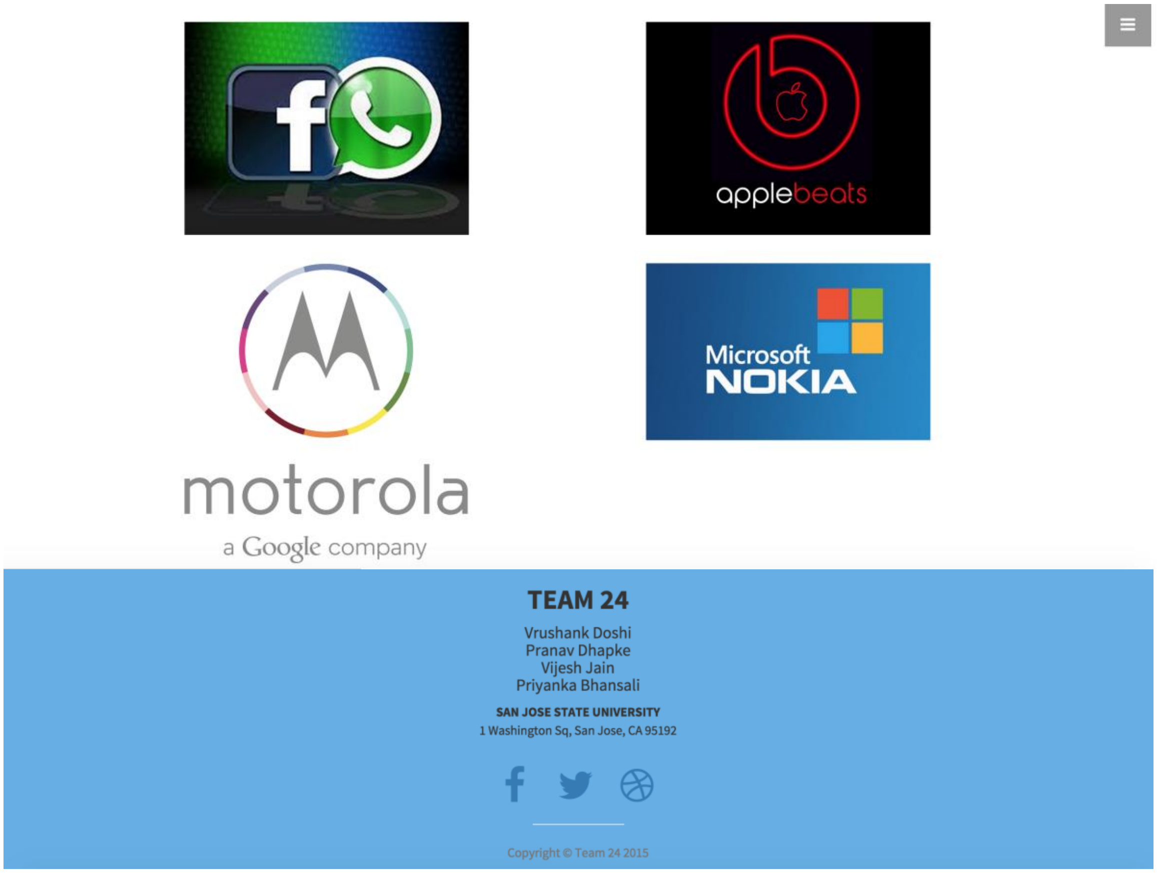


COMPANY ACQUISITION PREDICTION

The focus is on the future acquisition of companies in IT sector based on the financial data of company from past 3 years

It provides you a systematic analysis of the company's financial data and determine whether the company can acquire other company or can be acquired by other company in the future.





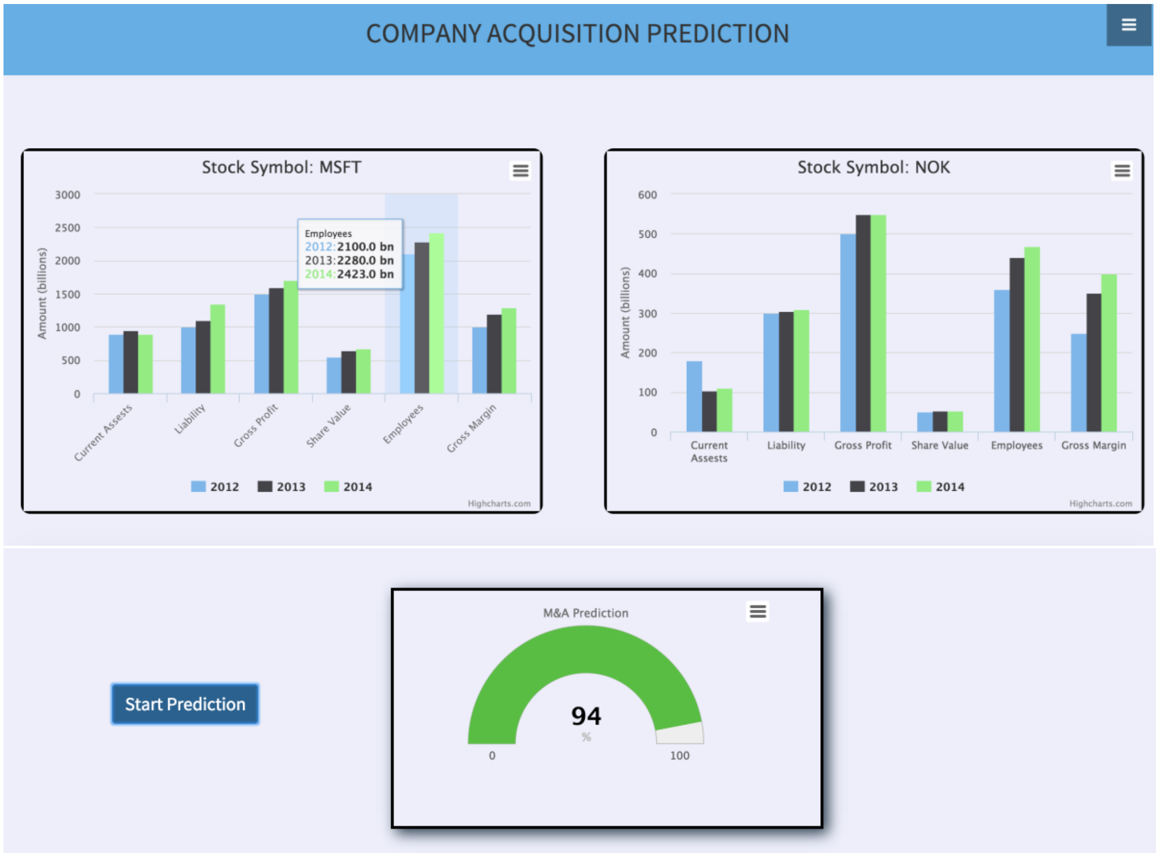
TEAM 24

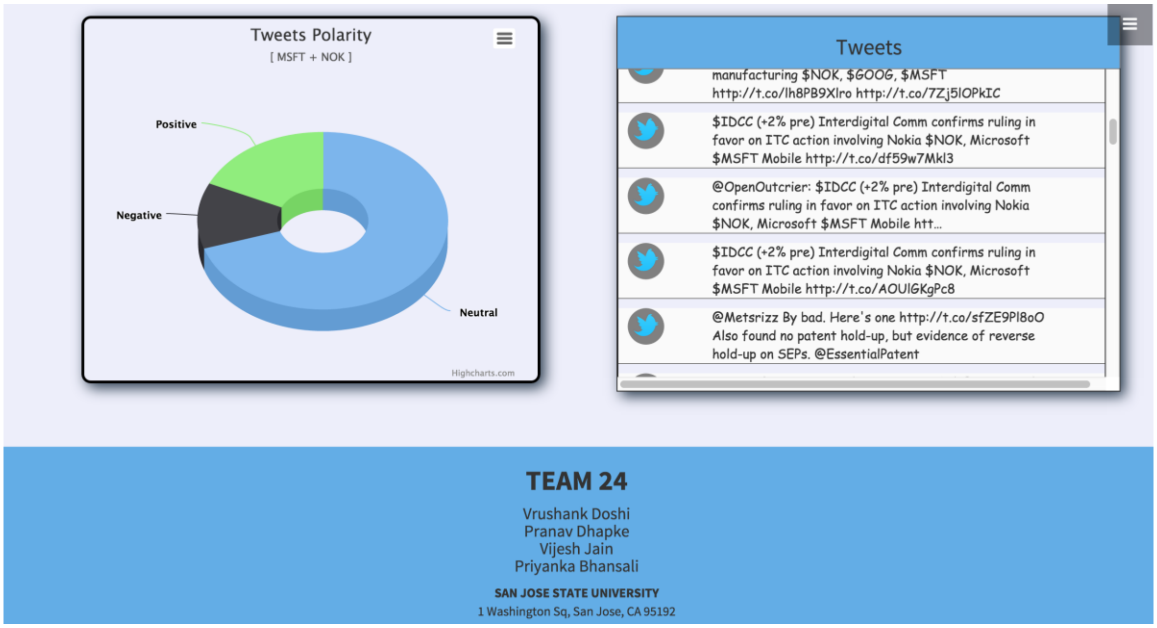
Vrushank Doshi
Pranav Dhapke
Vijesh Jain
Priyanka Bhansali

SAN JOSE STATE UNIVERSITY
1 Washington Sq, San Jose, CA 95192

Copyright © Team 24 2015

Statistics Page:





Tweets

manufacturing $NOK, $GOOG, $MSFT http://t.co/Ih8PB9Xlro http://t.co/7Zj5lOPkIC

$IDCC (+2% pre) Interdigital Comm confirms ruling in favor on ITC action involving Nokia $NOK, Microsoft $MSFT Mobile http://t.co/df59w7Mkl3

@OpenOutcrier: $IDCC (+2% pre) Interdigital Comm confirms ruling in favor on ITC action involving Nokia $NOK, Microsoft $MSFT Mobile htt...

$IDCC (+2% pre) Interdigital Comm confirms ruling in favor on ITC action involving Nokia $NOK, Microsoft $MSFT Mobile http://t.co/AOUlGKgPc8

@Metsrizz By bad. Here's one http://t.co/sfZE9Pl8oO Also found no patent hold-up, but evidence of reverse hold-up on SEPs. @EssentialPatent

TEAM 24

Vrushank Doshi
Pranav Dhapke
Vijesh Jain
Priyanka Bhansali

SAN JOSE STATE UNIVERSITY
1 Washington Sq, San Jose, CA 95192

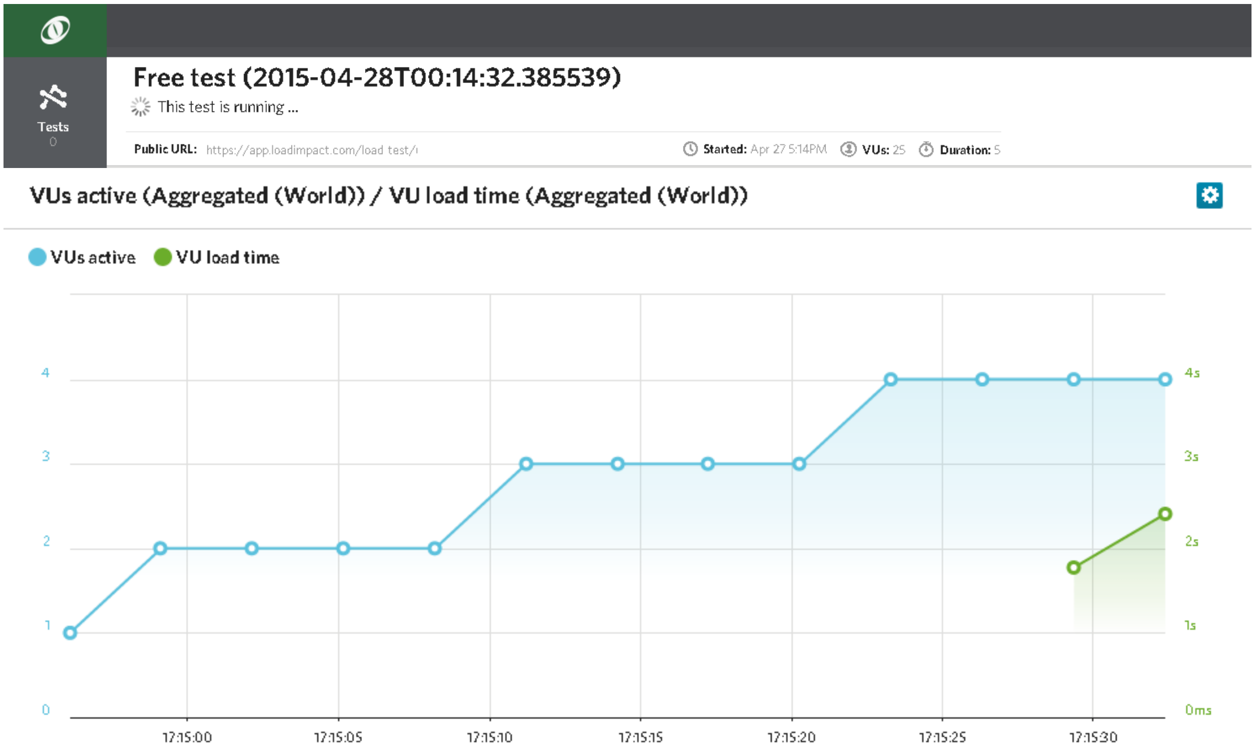
LOAD TESTING:

The load testing was conducted on our web application using 2 tools:

1. Load Impact tool
2. Pingdom

Load Impact tool:

In this tool we tried load testing our web application by supplying 25 virtual users on our application and checking its performance. We tested our application for 5 minutes by firing different number of users and monitoring its performance. Below the snapshots will give you the detail overview of the performance of our application with different number of users.

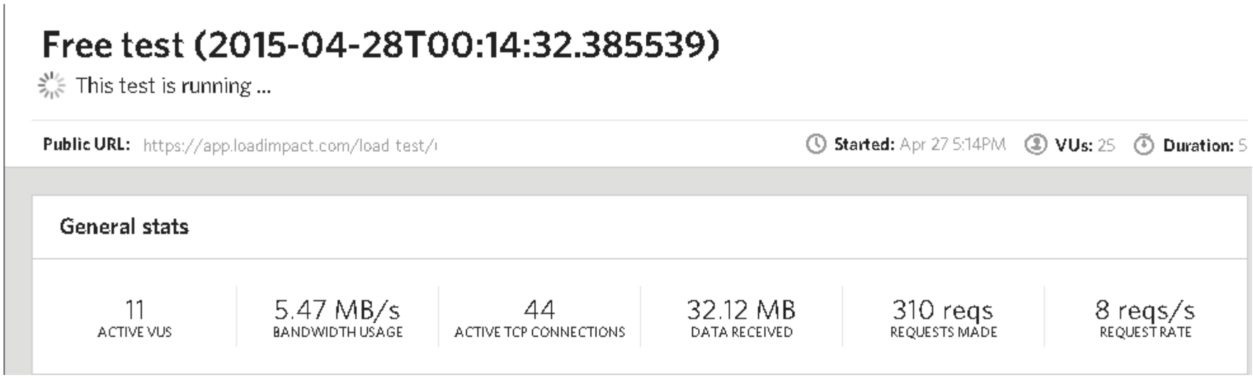


Free test (2015-04-28T00:14:32.385539)

This test is running ...

Public URL: https://app.loadimpact.com/load test/ **Started:** Apr 27 5:14PM **VUs:** 25 **Duration:** 5

VUs active (Aggregated (World)) / VU load time (Aggregated (World))



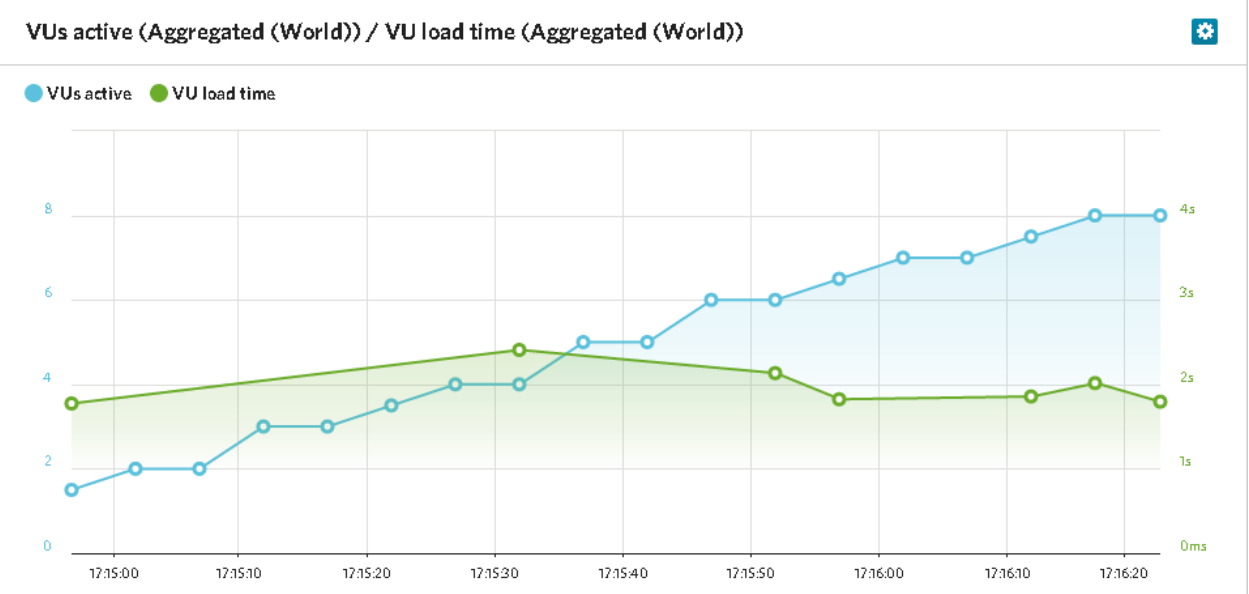
Free test (2015-04-28T00:14:32.385539)

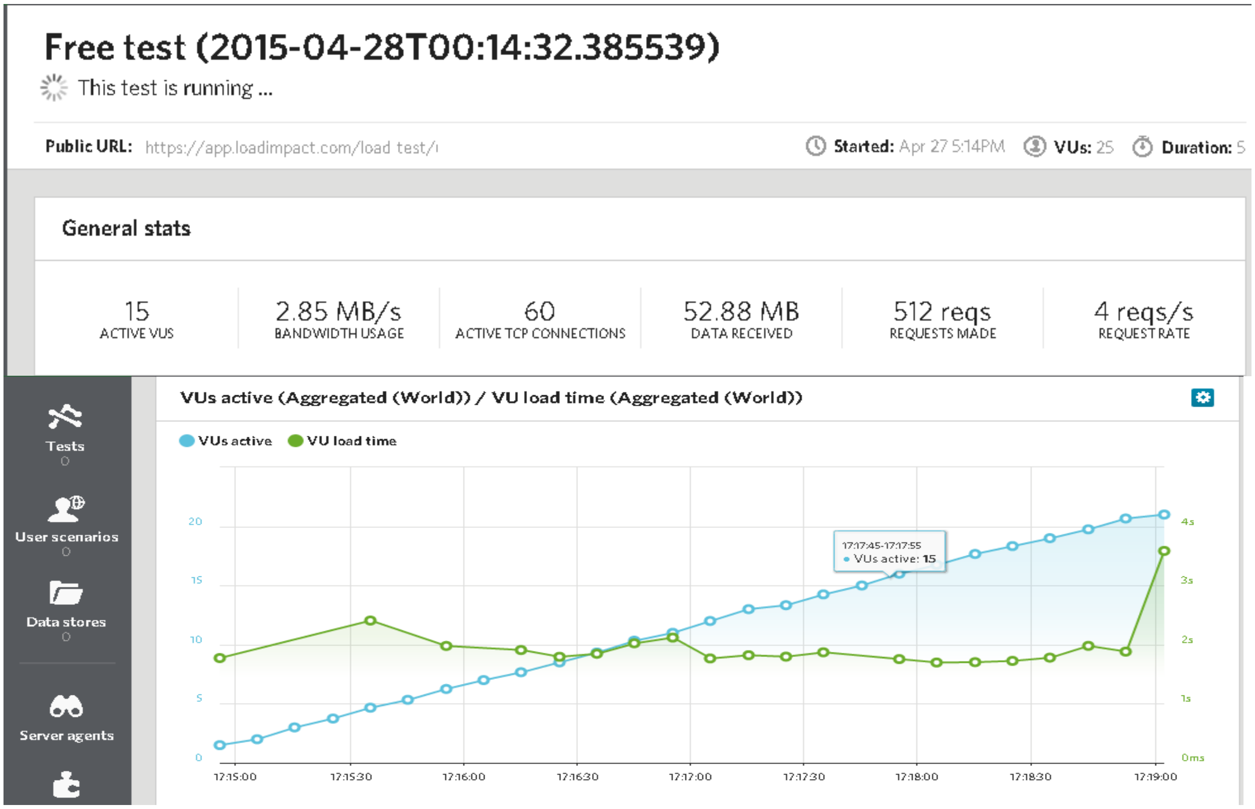
This test is running ...

Public URL: https://app.loadimpact.com/load test/ **Started:** Apr 27 5:14PM **VUs:** 25 **Duration:** 5

General stats

11	5.47 MB/s	44	32.12 MB	310 reqs	8 reqs/s
ACTIVE VUS	BANDWIDTH USAGE	ACTIVE TCP CONNECTIONS	DATA RECEIVED	REQUESTS MADE	REQUEST RATE





Free test (2015-04-28T00:14:32.385539)

This test is running ...

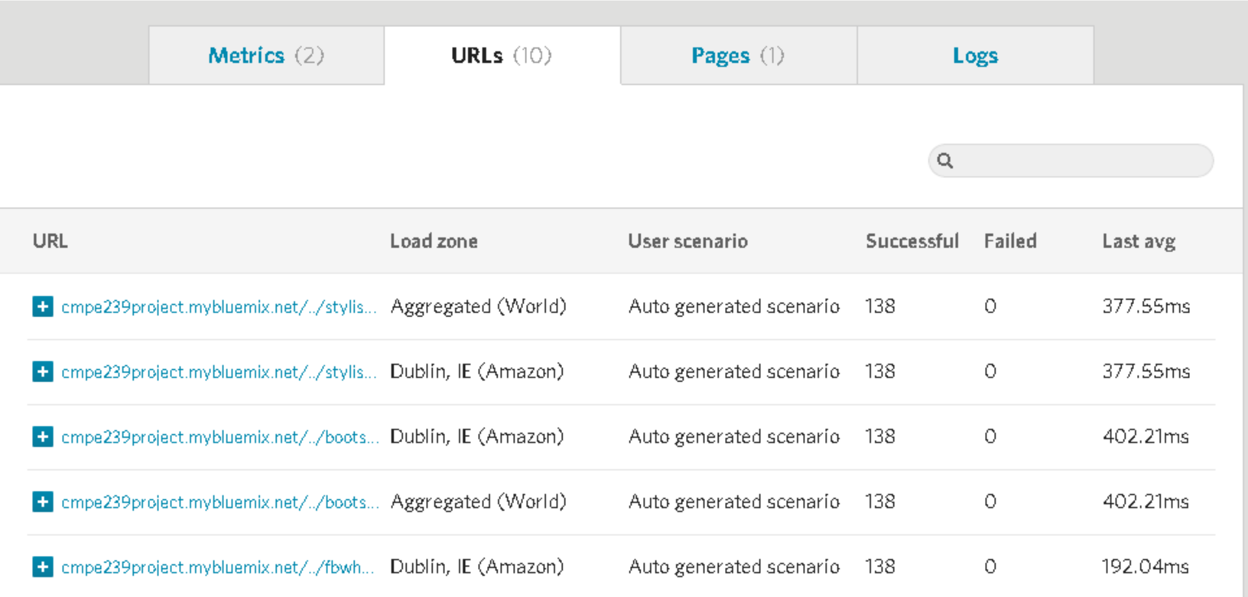
Public URL: https://app.loadimpact.com/load test/i **Started:** Apr 27 5:14PM **VUs:** 25 **Duration:** 5

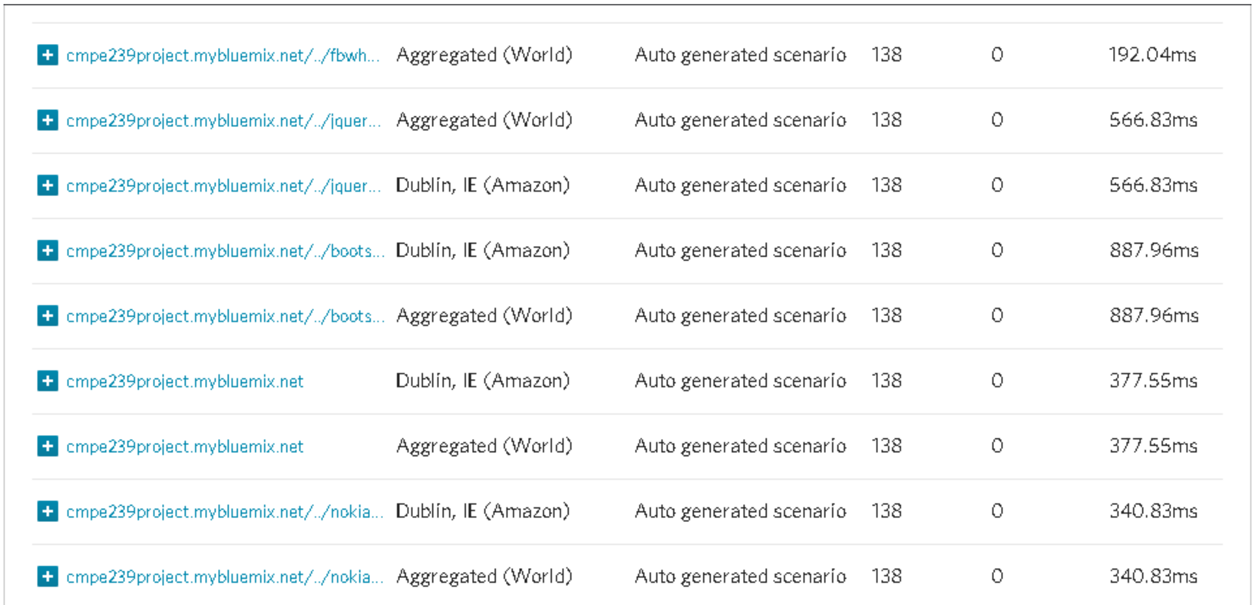
General stats

15 ACTIVE VUS	2.85 MB/s BANDWIDTH USAGE	60 ACTIVE TCP CONNECTIONS	52.88 MB DATA RECEIVED	512 reqs REQUESTS MADE	4 reqs/s REQUEST RATE

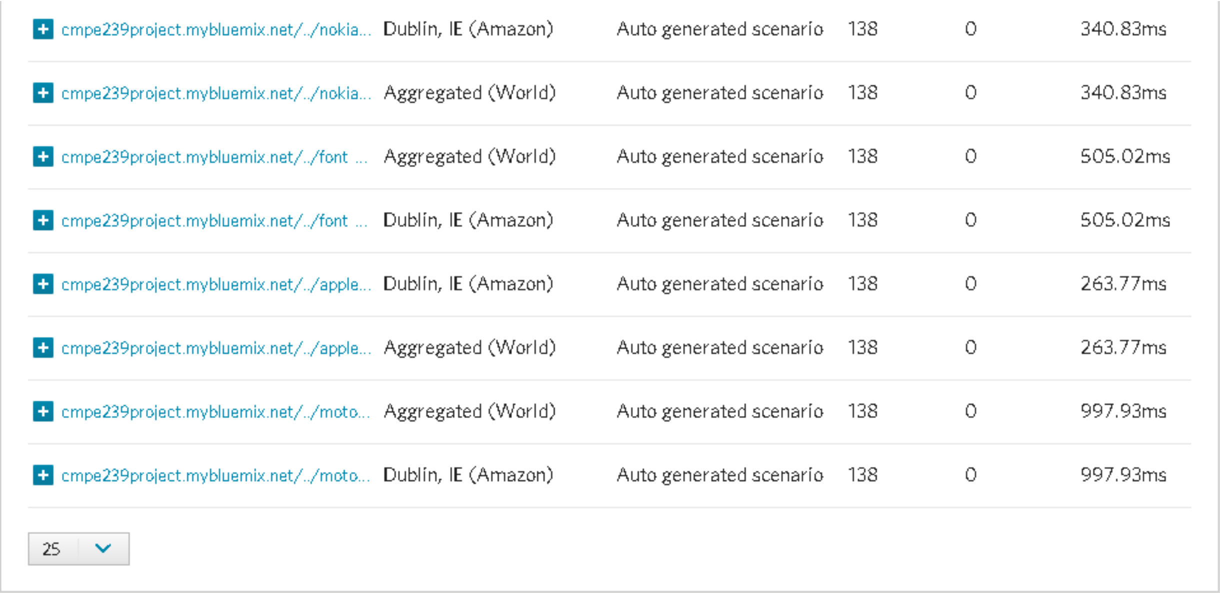
Following screenshot shows the result of 25 virtual machines.

These virtual machines were tested on auto generated scenarios. As you can see, there were 138 auto-generated scenarios and all are passed successfully.





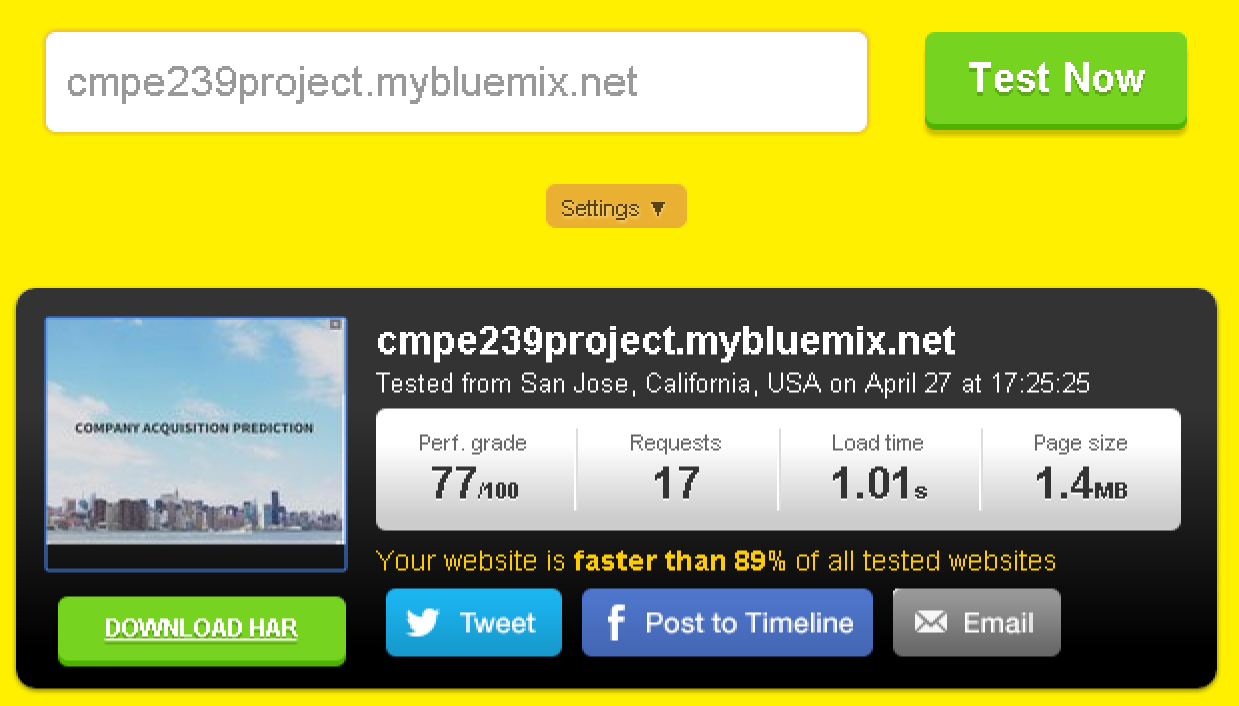
URL	Load zone	User scenario	Successful	Failed	Last avg
cmpe239project.mybluemix.net/../stylis...	Aggregated (World)	Auto generated scenario	138	0	377.55ms
cmpe239project.mybluemix.net/../stylis...	Dublin, IE (Amazon)	Auto generated scenario	138	0	377.55ms
cmpe239project.mybluemix.net/../boots...	Dublin, IE (Amazon)	Auto generated scenario	138	0	402.21ms
cmpe239project.mybluemix.net/../boots...	Aggregated (World)	Auto generated scenario	138	0	402.21ms
cmpe239project.mybluemix.net/../fbwh...	Dublin, IE (Amazon)	Auto generated scenario	138	0	192.04ms
cmpe239project.mybluemix.net/../fbwh...	Aggregated (World)	Auto generated scenario	138	0	192.04ms
cmpe239project.mybluemix.net/../jquer...	Aggregated (World)	Auto generated scenario	138	0	566.83ms
cmpe239project.mybluemix.net/../jquer...	Dublin, IE (Amazon)	Auto generated scenario	138	0	566.83ms
cmpe239project.mybluemix.net/../boots...	Dublin, IE (Amazon)	Auto generated scenario	138	0	887.96ms
cmpe239project.mybluemix.net/../boots...	Aggregated (World)	Auto generated scenario	138	0	887.96ms
cmpe239project.mybluemix.net	Dublin, IE (Amazon)	Auto generated scenario	138	0	377.55ms
cmpe239project.mybluemix.net	Aggregated (World)	Auto generated scenario	138	0	377.55ms
cmpe239project.mybluemix.net/../nokia...	Dublin, IE (Amazon)	Auto generated scenario	138	0	340.83ms
cmpe239project.mybluemix.net/../nokia...	Aggregated (World)	Auto generated scenario	138	0	340.83ms



+ cmpe239project.mybluemix.net/../nokia...	Dublin, IE (Amazon)	Auto generated scenario	138	0	340.83ms
+ cmpe239project.mybluemix.net/../nokia...	Aggregated (World)	Auto generated scenario	138	0	340.83ms
+ cmpe239project.mybluemix.net/../font ...	Aggregated (World)	Auto generated scenario	138	0	505.02ms
+ cmpe239project.mybluemix.net/../font ...	Dublin, IE (Amazon)	Auto generated scenario	138	0	505.02ms
+ cmpe239project.mybluemix.net/../apple...	Dublin, IE (Amazon)	Auto generated scenario	138	0	263.77ms
+ cmpe239project.mybluemix.net/../apple...	Aggregated (World)	Auto generated scenario	138	0	263.77ms
+ cmpe239project.mybluemix.net/../moto...	Aggregated (World)	Auto generated scenario	138	0	997.93ms
+ cmpe239project.mybluemix.net/../moto...	Dublin, IE (Amazon)	Auto generated scenario	138	0	997.93ms

Pingdom:

This is another load testing tool which checks every page and contents on the page for the time it takes to load on the application. As you can see in the following screen shots, it displays the performance of our application and the time it takes to load the page.



cmpe239project.mybluemix.net

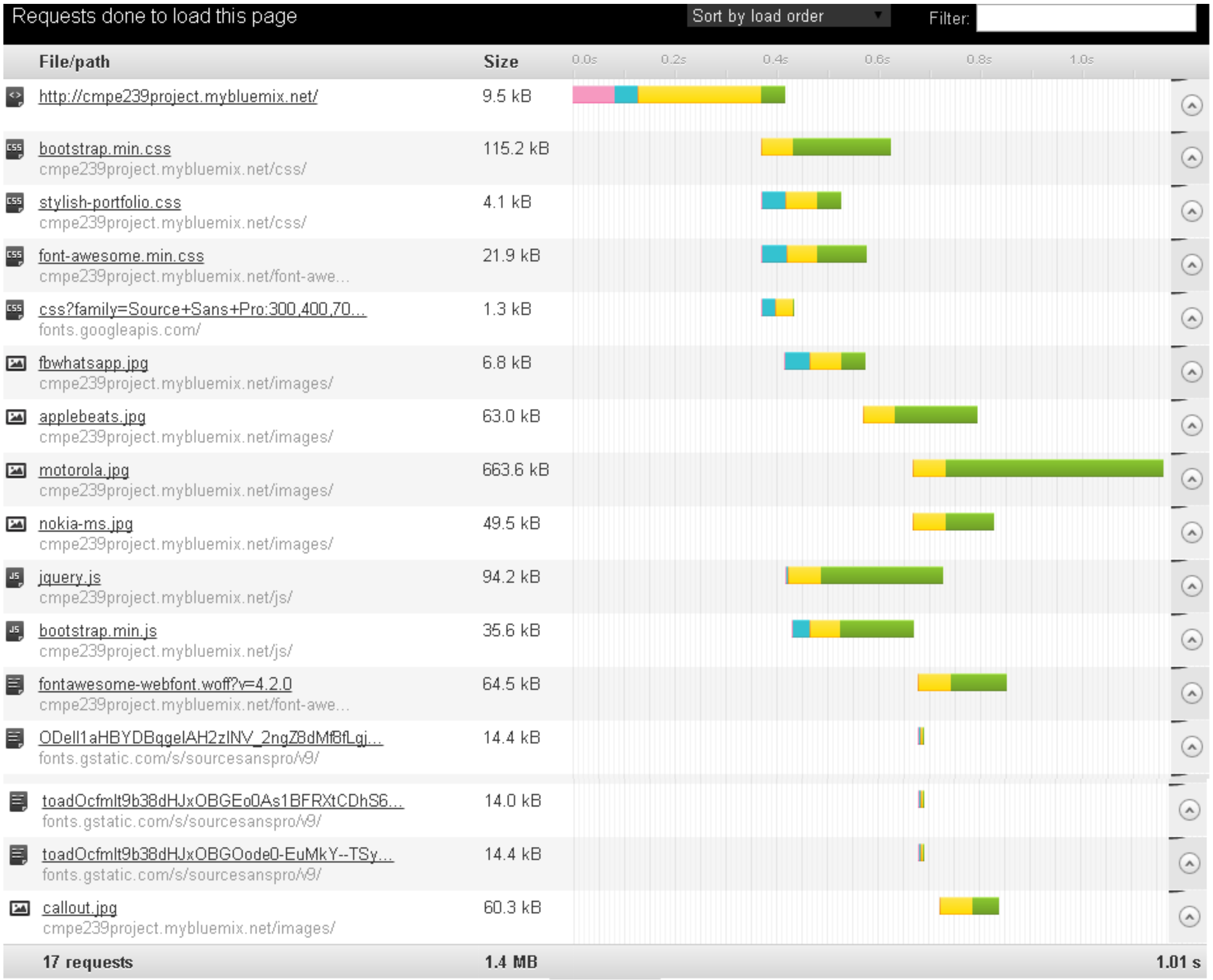
cmpe239project.mybluemix.net

Tested from San Jose, California, USA on April 27 at 17:25:25

Perf. grade	Requests	Load time	Page size
77/100	17	1.01s	1.4MB

Your website is **faster than 89%** of all tested websites

The following screen shot is a waterfall layout for each content on the application with the time it takes to load on the page. The total time taken to load the page is 1.01s.

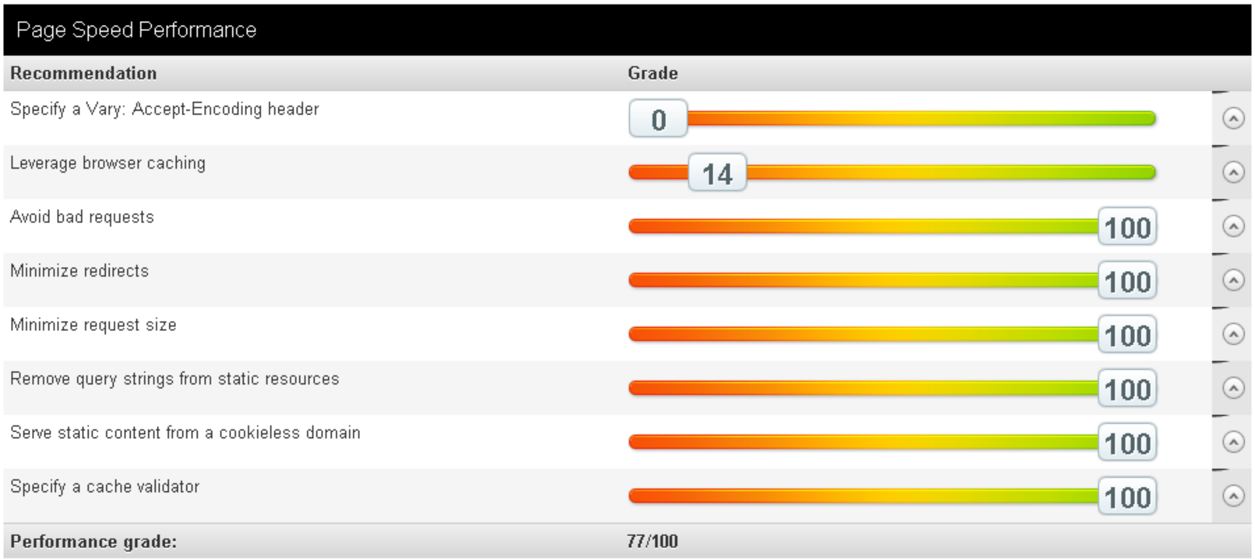


Requests done to load this page

File/path	Size
http://cmpe239project.mybluemix.net/	9.5 kB
bootstrap.min.css cmpe239project.mybluemix.net/css/	115.2 kB
stylish-portfolio.css cmpe239project.mybluemix.net/css/	4.1 kB
font-awesome.min.css cmpe239project.mybluemix.net/font-awe...	21.9 kB
css?family=Source+Sans+Pro:300,400,70... fonts.googleapis.com/	1.3 kB
fbwhatsapp.jpg cmpe239project.mybluemix.net/images/	6.8 kB
applebeats.jpg cmpe239project.mybluemix.net/images/	63.0 kB
motorola.jpg cmpe239project.mybluemix.net/images/	663.6 kB
nokia-ms.jpg cmpe239project.mybluemix.net/images/	49.5 kB
jquery.js cmpe239project.mybluemix.net/js/	94.2 kB
bootstrap.min.js cmpe239project.mybluemix.net/js/	35.6 kB
fontawesome-webfont.woff?v=4.2.0 cmpe239project.mybluemix.net/font-awe...	64.5 kB
ODell1aHBYDBqgeIAH2zINV_2ngZ8dMf8fLgj... fonts.gstatic.com/s/sourcesanspro/v9/	14.4 kB
toadOcfmlt9b38dHJxOBGEo0As1BFRXtCDhS6... fonts.gstatic.com/s/sourcesanspro/v9/	14.0 kB
toadOcfmlt9b38dHJxOBGOode0-EuMkY--TSy... fonts.gstatic.com/s/sourcesanspro/v9/	14.4 kB
callout.jpg cmpe239project.mybluemix.net/images/	60.3 kB
17 requests	**1.4 MB**

1.01 s

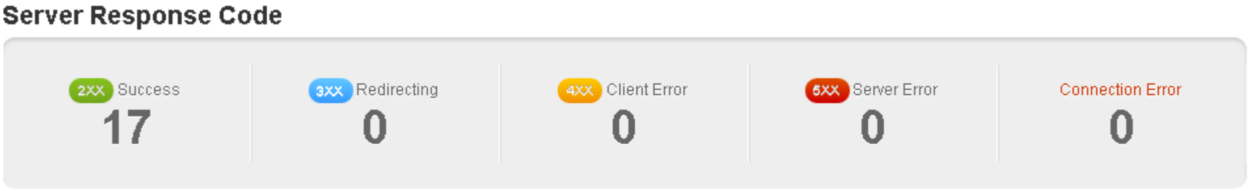
Page Speed performance:



Page Speed Performance

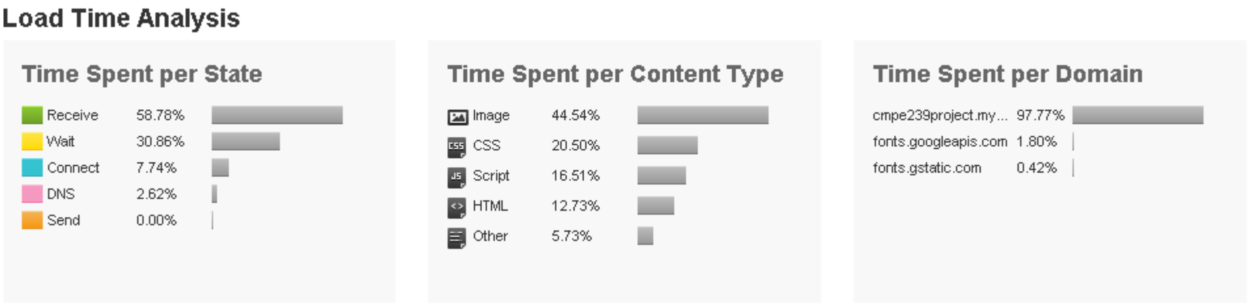
Recommendation	Grade
Specify a Vary: Accept-Encoding header	0
Leverage browser caching	14
Avoid bad requests	100
Minimize redirects	100
Minimize request size	100
Remove query strings from static resources	100
Serve static content from a cookieless domain	100
Specify a cache validator	100
Performance grade:	77/100

Server response analysis:



Server Response Code

2XX Success	3XX Redirecting	4XX Client Error	5XX Server Error	Connection Error
17	0	0	0	0



Load Time Analysis

Time Spent per State

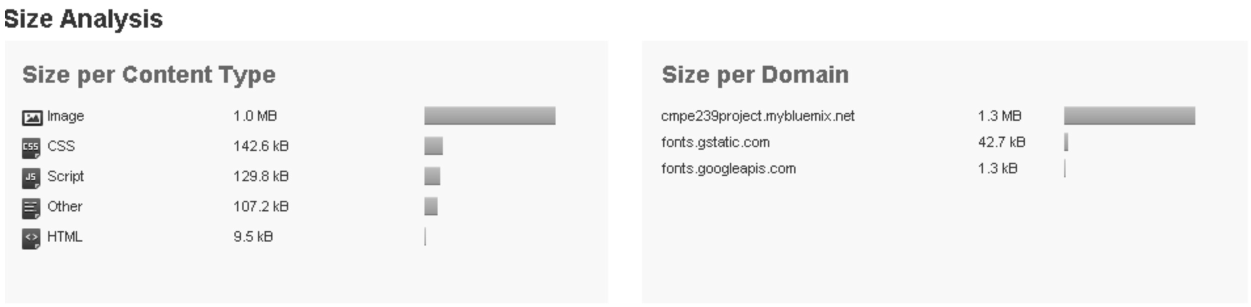
🟩 Receive	58.78%
🟨 Wait	30.86%
🟦 Connect	7.74%
🟪 DNS	2.62%
🟧 Send	0.00%

Time Spent per Content Type

Image	44.54%
CSS	20.50%
Script	16.51%
HTML	12.73%
Other	5.73%

Time Spent per Domain

cmpe239project.my...	97.77%
fonts.googleapis.com	1.80%
fonts.gstatic.com	0.42%



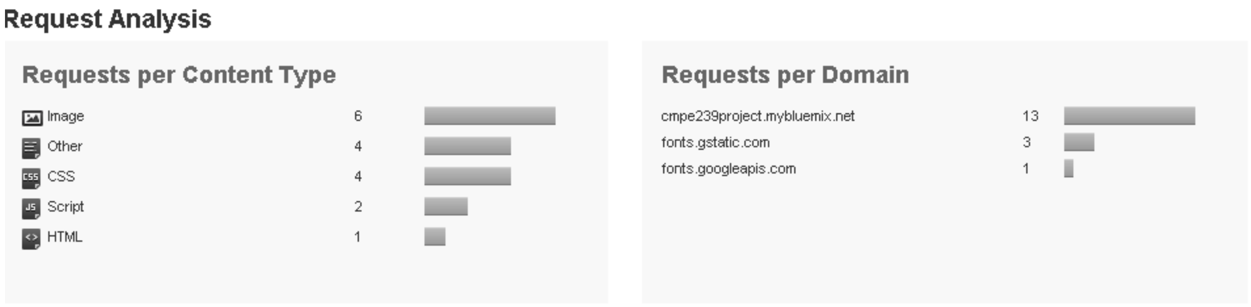
Size Analysis

Size per Content Type

Image	1.0 MB
CSS	142.6 kB
Script	129.8 kB
Other	107.2 kB
HTML	9.5 kB

Size per Domain

cmpe239project.mybluemix.net	1.3 MB
fonts.gstatic.com	42.7 kB
fonts.googleapis.com	1.3 kB



Request Analysis

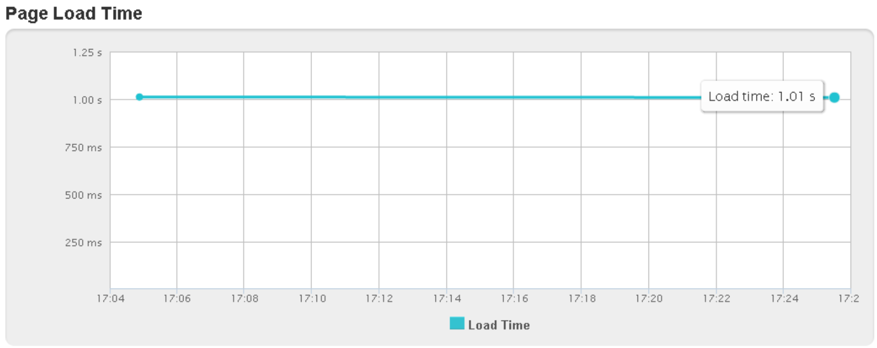
Requests per Content Type

Image	6
Other	4
CSS	4
Script	2
HTML	1

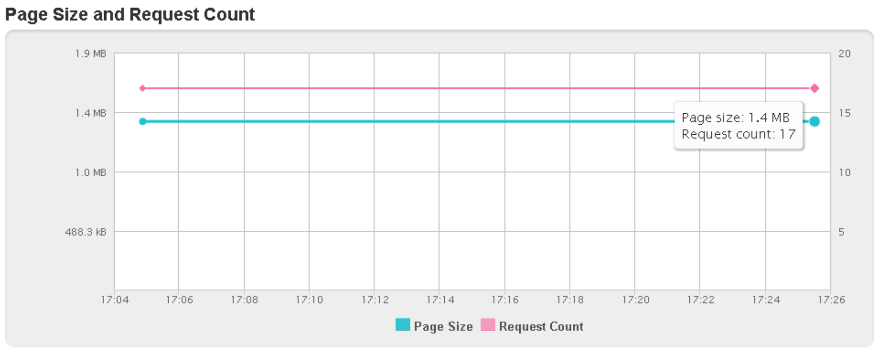
Requests per Domain

cmpe239project.mybluemix.net	13
fonts.gstatic.com	3
fonts.googleapis.com	1

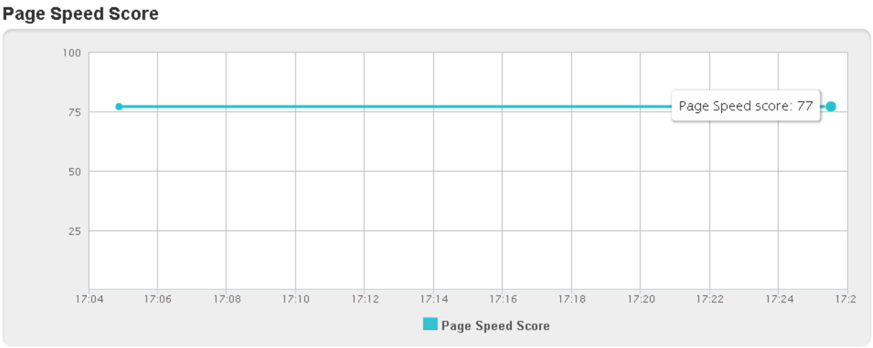
Page Load Time:

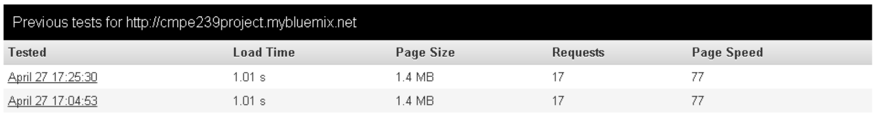


Page Size and Request Count:



Page Speed Score




Previous tests for http://cmpe239project.mybluemix.net

Tested	Load Time	Page Size	Requests	Page Speed
April 27 17:25:30	1.01 s	1.4 MB	17	77
April 27 17:04:53	1.01 s	1.4 MB	17	77

REFERENCES:

[1]. Wikipedia, 'Mergers and acquisitions', 2015. [Online]. Available: http://en.wikipedia.org/wiki/Mergers_and_acquisitions#Acquisition.

[2]. Finance.yahoo.com, 2015. [Online]. Available: http://finance.yahoo.com

[3]. Google.com, 'Google Finance: Stock market quotes, news, currency conversions & more', 2015. [Online].
Available: https://www.google.com/finance.

[4]. Kaggle.com, 'Kaggle: The Home of Data Science', 2015. [Online].
Available: http://www.kaggle.com.