June 17, 2018

Course: CIS570 – Business Intelligence

Name: Robert Palumbo

Assignment: Reading Discussions – Week5-Session2

Due Date: Sunday, June 17 @ 11:59pm

1. Question 1. What are some strategies to improve the ranking of web pages in search engine results?

We all are aware (or should be) if you run a business that has an Internet presence, having your business appear at the top of any Internet search rankings list and is hugely important to achieving business success. Companies that employ superior search engine optimization (SEO) techniques have a greater chance of being at the top of the search results than those who spend little time on SEO. There are many strategies that a business can use to improve their SEO and thereby enhance their Internet presence. A few of these techniques are listed here.

1. Publish Relevant Content.

This is clearly the most important SEO factor in my opinion. The intent here is that the presented content will draw users to your site increasing traffic flow and thereby giving providing more relevance to your site. In time, the increased flow will materialize in a higher ranking as the site’s importance rises over other sites of similar content.

When designing a web page, it should be designed to be as user-friendly and easy to navigate as possible. Poorly designed web pages turn users off and they will quickly leave your site and are detrimental to website rankings.

Likewise, keeping the content updated and current will go a long way to maintaining relevance of the site.

1. Use Metadata whenever possible.

Metadata is “*data about data”*. It is used to give context and even a description of a particular piece of data. In terms of web page design, there are several types of metadata that can and should be used including *Title Page, Description, and Keyword.*

Each type has its own importance in attracting users to your page. *Title* is very important as the contents of that metadata is displayed in the title bar of your main web page. *Keyword* metadata is important as well as these are data elements that are matched against user search criteria. The use of appropriate and relevant keywords is vital for a user search to be able to locate your website.

Remember, the more visits to your site the higher its ranking becomes!

1. Embedded Links

These are used to allow a user to navigate through different pages of your website and perhaps to navigate to other external sites. A good strategy for the use of embedded links to alternate pages is to encode (name) the link using specific keywords that are relevant to the site and to the requested page. A search engine will find it easier to optimize on specific keywords versus generic labels. An embedded link of *‘Click Here’* is not as useful as *‘Colorado Department Personnel’*.

1. Site Speed

Google will take into consideration the speed of a site. If the speed is too slow it will give it a lower ranking. Likewise, if a site is too slow users will quickly give up anyway and leave. It should become a general practice to frequently check the speed of the site and make improves wherever possible to increase access and processing speeds.

1. Determine your current sites ranking.

It is hard to improve your sites search ranking if you do not know what it is to begin with. Before you attempt any SEO enhancements, make sure you know where you rank to start. SEPRs.com can be used just for this purpose.

<https://www.mtu.edu/umc/services/digital/seo/>

<https://en.wikipedia.org/wiki/Metadata>

<https://neilpatel.com/blog/improve-google-rankings-without-getting-penalized/>

<https://serps.com/tools/rank-checker/>

<https://www.alexa.com/topsites>

Question 2. Visit Kaggle.com. Search for competitions (active and completed) that are based on social media and unstructured data. Briefly describe the competition including the sponsor, objective, reward, and datasets.

My search on Kaggle led me to the following competition: “Detecting Insults in Social Commentary.”

This objective of the competition was to predict whether or not a comment made by one party in a conversation should be considered insulting towards another party in that conversation.

The competition was sponsored by a company called *Impermium*. The company was founded in 2010 as cyber-security startup providing anti-spam (social media) and security protection for Internet websites. The company was very successful and in 2014 was acquired by Google.

This particular competition started on 9/18/2012 and ended on 9/21/2012. There were 50 *teams* that entered with a team size limited to 1 person (strange). Teams could submit 5 entries per day picking their top 5 for final judging.

The data as shown below is the *training* set that participants could use to train their model(s). As shown, the *comment* field is the actual social media text of a comment that was made during a conversation on the specified date and time as shown in the *Date* field.

The *Insult* column is the *target* variable and is used to *classify* the comment as either an *insult (1)* or *not-an-insult* (0). As we know, this is a *supervised classification* problem as the outcome is each occurrence in the training set is known ahead of time. Several technique could be used to classify the comments including word and phrase level analysis which attempts to match against known words and phrases that are known to be insulting. More sophisticated techniques could be used to determine the *context* of the comment to classify as well.

Each submission would be run against a *validation* dataset. The model would predict a 0 or 1 for each included comment which would be recorded in the *Insult* column. The final output for a model run would be submitted and the prediction accuracy result would be calculated and returned back to the participant, so they could determine how well their model was performing and attempt to make corrections to improve performance. A final *test* data set was used for assessing the performance for each of the solution models submitted by each team for fairness in judging. The winner of the competition was the team that submitted the model that achieved the highest predication accuracy percentage.

According to the final results, the 1st place team scored 84.248%, 2nd place team scored 83.977%, and the 3rd place team scored 83.687%. These scores are actually very good considering the difficulty in discerning meaning in the processing of textual data.

The competition awarded 3 cash prizes: 1st place - $7,000, 2nd place $2,500, 3rd place $500. The sponsor also used this competition as a recruiting opportunity for a *Principal Data Engineer.* Top performers of the competition were selected to interview for the position. Pretty sweet way to land a job.

Sample training data:

|  |  |  |
| --- | --- | --- |
| Insult | Date | Comment |
| 0 | 20120528192215Z | "i really don't understand your point.\xa0 It seems that you are mixing apples and oranges." |
| 0 |  | "@barrettmarson Huh? Her income was $21,912. Her spending was $5,470. My math puts that at 24.9%." |
| 1 | 20120611090207Z | "FOR SOME REASON U SOUND RETARDED. LOL. DAMN. WHERE U BEEN NEGRO" |
| 1 | 20120320162532Z | "You with the 'racist' screen name\n\nYou are a PieceOfSh$$.........." |
| 0 |  | "Oh! & cheating as well...1966 cup was a cheat fest, & you still dare to post that comment..shame on you lol" |
| 1 | 20120320075347Z | "your such a dickhead..." |
| 1 | 20120320203947Z | "Your a retard go post your head up your #%&\*" |
| 0 | 20120620055647Z | "@sdnoriko LOL as if Capcom listens. Watch them f&ck this up too. I hope I am wrong. I love Okami." |
| 1 | 20120515132156Z | "Allinit123, your\xa0hypocrisy\xa0is sickening, you ignorant loser." |
| 0 | 20120620011710Z | "Fast & furious didn't work out like they planned did it?.............lololololololol" |
| 0 | 20120530111709Z | "mitt isn't the best candidate for president.......but, going against the worst\xa0candidate\xa0ever should help him tremendously. \xa0LOL\xa0\n\nobama is one and done......it could happen no other way." |
| 1 | 20120620161958Z | "I can't believe the stupid people on this site that have no idea what is going on.\xa0 You fools.\xa0 You are all useful idiots.\xa0 They are hiding the truth because you fools who protect him would look like the useful idiots you are." |
| 1 | 20120609163743Z | "f#cking behave then you prick!" |
| 0 | 20120619094846Z | "The only "other" relevant fact is that HE IS DESTROYING AMERICA!" |
| 0 | 20120609175446Z | "You are a liar." |
| 1 | 20120620075045Z | "Is that it?? your sum total to any conversation is first abuse and then calling them a troll.. You are a typical socialist idiot. Keep up your non existent class warfare, it is as pointless as you." |

Sample Verification Dataset:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | Insult | Date | Comment | Usage |  |
| 1 |  | 20120603163526Z | "like this if you are a tribe fan" | PrivateTest | |
| 2 |  | 20120531215447Z | "you're idiot......................." | PrivateTest | |
| 3 |  | 20120823164228Z | "I am a woman Babs, and the only "war on women" I see is coming from jackazzes like you.\xa0 I don't need your protection or your ignorant rhetoric masquerading as representing my best interests.\xa0 " | PrivateTest | |
| 4 |  | 20120826010752Z | "WOW & YOU BENEFITTED SO MANY WINS THIS YEAR FROM HIS BAT NICE STUPID" | PrivateTest | |
| 5 |  | 20120602223825Z | "haha green me red you now loser whos winning now moron" | PrivateTest | |

<https://www.kaggle.com/c/detecting-insults-in-social-commentary>

<https://en.wikipedia.org/wiki/Impermium>