Foundation for Critical Thinking

P.O. BOX 220 Dillon Beach, CA. 94929
Tollfree: 800.833.3645 - Fax: 707.878.9111
cct@criticalthinking.org

Print this Page

The Logic of Analyzing a Problem:

Linda Elder - Foundation for Critical Thinking

The problem you are analyzing: Popula

Created Friday Nov 10 2017, 6:12 AM
Edit/Return to your Problem

The logic of the question you have reasoned through can be found here. Beside each of your answers, you will find checkpoints, or guidelines for making sure you have reasoned well through each element. If you are unclear as to how to interpret any of these "checkpoints," read more deeply about the elements of reasoning and intellectual standards.

Statement of the problem:

When the World Wide Web was first invented in 1989, it was a very rudimentary version of what it is now. Since then, the Web has expanded to unconceivable heights, one vertical of which is online news. Due to this, prediction of the popularity of online news is becoming a trendy topic. But, the popularity of the news castings depends on “Entertainment/Lifestyle ” and “Word ratings”, two factors which are not easy to predict. Thus, it is difficult to predict user behavior to know if the published news will become popular or not, both in general and to a specific audience. In this ocean that is online news, only a few websites will become popular. This can be seen in the disproportionality between the number of online news to the popular ones. Our main aim is to predict the popularity of online news articles before their publication, and then provide corrective action to the authors by showing the number of article features which need to be altered in order to enhance the popularity of their news.

- The key **question** at issue is...

One of the major issues which we predict will arise with this project is to identify the subset of news features which would enhance the popularity of a news article. Questions that will arise are: 1. How can we relate the popularity of a news article based on features like “rate_positive_words”, “rate_negative_words”, “token_title”, “token_content”, “num_videos”, and “num_images”? 2. Do we have enough un-biased data sources regarding online news publications such that we can build more a more stable system without losing data integrity? 3. Which attributes will have the most impact on popularity prediction? Is our dataset reliable and able to provide such attributes? 4. Is the dataset retrieved from legitimate and reliable source? 5. How can we pre-process and clean our data set so that it can fit for the data-modeling process and obtain the best possible prediction results?

Clarify the question: Have I stated the question as clearly and precisely as possible? Have I stated the question so as to detail the complexities in it? Does my formulation of the question give me [or others] some unjust advantage?

- The **purpose** of reasoning through the issue is...

We plan to design a web based decision supporting system where the user inputs some related to the online news or articles to know the popularity level of that particular news. Most authors would be worried about the features of an article which would indirectly affect the popularity of their said article. Our system will guide such authors, by predicting the popularity of their articles in terms of a percentage and then providing them a list of attributes which need to be changed in order to enhance the popularity of their articles. Also, authors/publishers are eager to know the popularity of their articles which have already been published as well. Our decision support system will provide such authors/publishers with suggestions to improve the news features which would grasp much more attention of users, and feedback on what went wrong in an old article which need not gain many views. People tend to view the popular news based on ratings and comments. Popularity reflects the interest of the people in such news casting. Ranking the news articles according to their popularity will improve the chances of that news being visited more by users.

Check your purpose: Have I stated the question as clearly and precisely as possible? Have I stated the question so as to detail the complexities in it? Does my formulation of the question give me [or others] some unjust advantage?

- The **information** I need to use in answering this question is...

In order to predict the probability of popularity, we would require an unbiased data set with the required news feature to help design a more robust and precise predictive data model. Data will be unbiased when there is equal proportion of instances with “High”, “Medium” and “Low” as values of the target class which shows popularity of an article. We have found one such legitimate data source, the UCI machine learning community, from where we can obtain a dataset regarding news articles which were published on the “Mashable” forum. It has around 39,644

Assess your information: Have I included all the important relevant information? Have I left out any relevant information I would rather not have to consider? Have I checked to see that my information is accurate?

instances and 60 attributes. As our output label “Shares” is a numeric attribute having 1167 distinct values , we decided to discretize the output label into 3 distinct values (High, Medium, Low) using discretization process. Distribution of instances among 3 distinct values is as follows : High – 6670 (28.96%) , Low – 8856(38.56%) , Medium - 7720 (33.21%). Also, we are planning to use one of the feature optimization algorithm in order to find out the important features of news which could help author to enhance popularity of article by making few changes in those extracted features.

- The **assumptions** I am making when reasoning through this issue are...

There are certain assumptions we will make while solving the problem. They are given below: 1. The popularity of the news will mainly depend on some of the attributes like “rate_positive_words”, “rate_negative_words”, “token_title”, “token_content”, “num_videos”, and “num_images”. 2. The popularity prediction will increase the quality of the news and also contribute to the growth of news sources 3. An online news popularity prediction model can be created with the existing data-mining techniques and software 4. Publishers/authors would provide legitimate attribute values of the article to our proposed model. 5. Our proposed model will be supportive to both users as well as publishers in terms of knowing the popularity of news and taking appropriate actions.

Identify assumptions: Do I recognize the beliefs I have taken for granted in reasoning through this issue? In other words, have I uncovered the important assumptions guiding my thinking in this situation? Have I questioned assumptions that may not be based in sound reasoning? Am I holding onto assumptions that cannot be justified based on the evidence?

- The primary **concepts** guiding my reasoning about this issue are...

One of the idea which we could use while developing the news popularity predicting model is making use of “rate_positive_words”, “rate_negative_words”, “token_title”, “token_content”, “num_videos”, and “num_images” as few of the important factors in deciding the popularity of news. “Popularity” of each news would be our target attribute which would be predicted based on number of other important features of news articles. We would be using Discretization initially to group the class label into 3 distinct values (Low/Medium/High). This training set is then used to train the model using classification algorithm. Before building the classification model we would preprocess the data using data-preprocessing steps like aggregation, Removal of irrelevant features and Feature subset selection to remove all the irrelevant attributes. Later, optimization will be performed on output data of trained model in order to select the required news features which need to be altered for popularity enhancement of news. One of the concern as stated earlier while predicting popularity might be geographical location at which that particular news has been published. Say, news1 published at location 1 was predicted as popular. But, it may not be the case that the same news1 published at another location 2 would get popular there as well. Our model may term “News1” as “Popular” at location 2 as well, though it may not become popular in future. This issue we can consider as the main issue of our problem which notifies geographical location might affect the popularity of the News.

Understand the concepts: Have I identified the main concepts guiding my thinking through this issue? Am I even clear about what a concept is and the role concepts play in human thinking? Have I distorted some idea to fit my vested interest? Do any of my concepts need to be questioned? Have I thought deeply about the concepts I am using? Am I using concepts superficially?

- Some important **implications** of reasoning through this problem well or poorly are...

Online news are mainly enjoyed by the mobile users and are massively spready through the online social platforms resulting in the increased interest in discovering the news that will become popular among readers. The content popularity prediction has direct implications in the development of new services for online news publishers/authors before distributing it to the readers. This model might be hailed as an adequate solution for online news popularity prediction and can be considered as a valuable source for news ranking. But depending completely upon this model may also result in certain unpredictable circumstances because this model will predict based on the history of the similar kind of news worldwide but the published news in that certain area might not get the expected attention of the readers as it might be of no interest to them. Hence, we got to keep in mind what news is being published and what area is it going to be published and does the news attract the readers in that area also does this kind of news has any positive or negative results.

Recognize the implications: Have I traced out the important implications in this situation? Have I thought through the potential negative as well as positive implications connected with this issue? Are there implications I would rather not face, and so I am refusing to consider them? Have I anticipated the implications of the obvious implications, and then the implications of those implications, and the implications of those implications, and so on, as well?

- The **point of view** from which I am looking at this situation is...

This model will act as a reference to help the news publishers/authors to predict the popularity level of the news being published. It does not give ultimate match for the news which is yet to be published but will give certain predictions based on the history of those same kind of news. Another view is that this model will the help the authors/publishers to take certain precautions before publishing the news online and will also help to

Consider your Point of View: Have I considered all important viewpoints relevant to this situation before formulating my viewpoint? Have I inadvertently distorted some other viewpoints in order to maintain my viewpoint? Have I articulated and considered other ways of looking at the situation, in good faith, before coming to conclusions about how to think and how to act?)

rectify the published news inorder to attract more popularity among the readers. From the user point of view, People tend to view the popular news based on ratings and comments. Popularity reflects the interest of the people in a news. Ranking the news articles according to their popularity will improve the chances of that news being visited more by users. So, we will have to take the above considerations in order to make our prediction model successful.

- In reasoning through this issue, the main **inferences** I have made and **conclusion(s)** I have come to, in terms of how to deal with the issue are…

Prediction of news popularity will certainly be increasing the degree of publication of more polished and useful news around the world. News popularity prediction will allow publishers/authors to enhance their news and make it more user beneficial. This prediction model will always urge the authors to refine their contents periodically in order to sustain within media world. Lower the user views on particular news, lower will be its sharing count and thus lower will be its popularity which could badly impact the publishers in terms or economic and social aspects. Thus, author would periodically check the popularity prediction via our model and would frequently make changes in article in order to withstand in market. As far as users are concerned, our news popularity model would allow them to have easy access to most important and legitimate news based on its popularity. Users will always prefer to visit top ranked news based on popularity metrics. Thus, we could say, such decision predictive system would make larger impact on media world while forcing them to publish much more polished and useful news around the global online world.

Analyze your interpretations: Am I clear about the inferences I have made in coming to this conclusion? Have I considered the fact that any or all of my inferences may be more or less sound? Do my inferences clearly follow from the evidence, or have I failed to consider important information in coming to my conclusions? Can I articulate the assumptions that have led to my inferences? Do I have a vested interest in coming to a particular conclusion, and if so, has this clouded my judgment?

Summary:

While reasoning through the logic of this problem, I have gained the following insights and come to the following decisions.

With the development of Web, prediction of online news popularity has been a growing interest. This Decision Support System will first extract a wide range of features both prior to the publication and after publication of a news article in order to enhance future popularity. From UCI publication, we have all the features extracted for the prediction of published articles. The decision support system for prior popularity prediction will extract the features from already available dataset by removing few attributes related the future publications. On the large dataset containing around 40,000 news publications we will be performing certain classification models in order to achieve the highest accuracy. We need to work on optimization techniques in order to extract the important features from the news article which would help in enhance the popularity of news articles before and after publication.
From the overall logic of problem, we have gained insight that predicting popularity would be a difficult task to solve, and hence we need to pay more attention to the same. Another major issue will be to predict the popularity of a news article based on the location of the published article.
Future scope of this Decision Support System will be to enhance the existing model so that it can predict the popularity of a news article based on location of publication.

Save Conclusions / Finish Problem

If you are unclear as to how to interpret any of these "checkpoints," read more deeply about the elements of reasoning and intellectual standards. We suggest the following materials:

- The Thinker's Guide to Analytic Thinking
- Critical Thinking: Tools for Taking Charge of Your Professional and Personal Life
- Critical Thinking: Tools for Taking Charge of Your Learning and Your Life