MAKERERE****UNIVERSITY

**COLLEGE OF COMPUTING AND INFORMATION SCIENCES.**

**BACHELOR OF SCIENCE IN SOFTWARE ENGINEERING**

**BSE2301 RECESS**

SUPERVISOR: MBABAZI ISAAC

**GROUP 8**

|  |  |  |
| --- | --- | --- |
| Name | Student Number | Registration number |
| Kimwero Dickson | 216000635 | 16/U/441 |
| Lumala David | 216003711 | 16/U/6679/eve |

**Concept paper on the effect of writing quality on datafiniti’s product database**

**Introduction**

To convince someone that something is actually good, your review must hold a lot of content and one of those are the good writing skills. Its very easy to rate a review with no spelling errors, good punctuation and many more good writing skills aspects otherwise the reviewer on finding a mistake may just move on to the next review without completing the previous one.

Everyday a person is misguided into buying a wrong product or profits from getting the right product from platforms like data Infiniti that provide access to product reviews due to poor communication from reviews some of which arise from poor writing quality.

Information written with all the quality is very useful because when you misspell a name, you may instead refer to a different product and rate it poorly yet it’s the best product and you meant a different one and vice versa. This leaves new entrants to online business hate it yet they were misled by errors.

We are interested in investigating this problem to improve the accessibility of reviewers with little knowledge about what they want but still being able to find what they are looking for easily. Its not an easy thing to satisfy everyone but it’s better to leave the vast majority satisfied and with the correct information.

**Problem**

Writing quality has greatly affected the reviews on the internet starting from wrong spellings to improper punctuation and many more. This is a vast problem and since you intend to communicate to people with little know how, you have to aim at making them understand whatever it is about a product. This is one of the biggest problems affecting reviews and we would like to manipulate the data given and present its vast effects.

With the vast amount of data that data Infiniti has, my team made of Kimwero Dickson and Lumala David we aim to manipulate the data and come up with the impact on how writing quality and mistakes affect the reviewers like in the following ways;

**Our main objectives are to investigate the following;**

**General objectives**

We aim to find out if the reviewers use punctuation correctly. Its very difficult to properly understand what a person meant if they have an error in punctuation. If a sentence is wrongly punctuated or if the proper punctuation is missing, the message may not be communicated as intended. We shall sieve the properly punctuated from those that are not.

We aim to find out the number of spelling errors that are made by those making reviews. For proper communication, what you type must be correct otherwise it will result into a lot of mistakes by the reviewers often misleading them for example some words if you miss out a letter then it’s a different word hence we aim to investigate that more.

People rate different things differently according to how they feel and their experience for example here in Uganda we are new to Uber but those that have used it are asked to rate the driver and they give stars according to their experience. It’s the same thing here and we aim to find out the ratings based on many factors like length of review, spelling errors etc.

**Specific objectives**

We shall also aim to find out if the number of spelling errors differ by rating. People rate differently according to how they feel, based 5 stars if one is not satisfied with the service they may provide one star and if they are satisfied they may give 5 stars. We aim to find out the different ratings on the number of spelling errors if they exist.

Some products are similar and can be used as substitute goods whereas some are completely different but all in all we aim to find out the star ratings across products, to see which product is better if there are substitute goods and see which one is more used and liked and to find out across a variety of products which is the best product still based on the ratings given by the purchasers.

Some reviews can be short while others can be long, we aim to find out which ones are rated better. Some people like long reviews where they can get full details about something while others prefer short summaries of everything hence we will manipulate the data and provide a report on the ratings across the length of the review to find out which ones perform better.

Like I communicated earlier reviews can be short or long and based on our current dataset we would like to investigate and find out how long a typical review can be by checking the most reviews which may involve finding an average or to see which ones are most commented about to which will communicate the best length for a review.

We also aim to find out the number of occurrences of spelling errors per unit time. How often do these spelling errors occur and when mostly do people make them?

It is very easy to forget some punctuation marks like a full stop at the end of every sentence. We would like find out the number of reviews that don’t end the sentences with the proper punctuation.

Examining all the data fully, we want to provide you with a proportion of reviews with spelling errors so that you can easily know that out of these number of reviews, this number have spelling errors.

**Methodology**

We don’t have all the time in the world so we aim at doing this in the least time possible and according to our objectives that is what we want to achieve, we chose to use two development methodologies and that is Rapid Application development where we shall need deliverables because we need the results to show progress and because of the time factor

We also aim to use joint application development because we may not meet every time and we may be developing and finding every thing offsite and the group work involved we shall be able to deliver in time

We are planning to use a lot of analysis to get the data for example coupled with our skills in R programming we are planning to produce results in the most understandable way. This is going to be like an example of unsupervised learning where our main aim is to find out hidden patterns in the data without trying to make any predictions and most of our interactons are going to be on github where we sent you an invitation already

We are planning to use the least time possible and that is 6 weeks and we shall present to you all the details of the whole project.

**Overview**

Writing quality is key to the internet because of the vast of information that exists, you are communicating to very many people who may be well versed with the language and they may be the ones to read the reviews because they understand the language very well. You have to be careful when writing when communicating with people.

This data shows on average the reviews followed and it shows that most of the ones with errors are neglected not because they are too understandable but because people don’t have time to waste on understanding whatever went wrong for your review. Hence you have to be careful when writing one.

For our budget, we have all the requirements we need and we are only going to present to you the final project.

By the end of the project, you will be able to know how each writing mistake be it a punctuation affects a review someone puts up. A summary including graphs for easy understanding will be included and all in all we hope for the best and people to take not while writing the reviews.

**References**

* [Gary\_B.\_Shelly, \_Harry\_J.\_Rosenblatt] \_Systems\_Analysis(BookFi.org) [1]