Cover Letter

Dear Mr. Kuan:

We are Group 10 professional with strength in the aspects of project management, user interface design and programming ability. The objective of our project is to deliver the IT solution of helpful review identifier to benefit the consumers with the plugin browser extension in which we take machine learning algorithms, and design database including analysis of survey and research to deliver the prototype of an identifier.

Some highlights of our career include: learning of project management, system programming, user interface design, artificial intelligence, database, data structures and computer algorithms. That's knowledge support us to design the prototype and strategize the project properly.

Attached is the final report of our project team. We look forward to further discussing the approval of the proposal. Thank you for your time.

Yours Faithfully,

Group 10

The date of the final report submitted: 29 October 2019

Table of Contents

1. Executive Summary

Dr. Kevin Kuan is the client of helpful review identifier project, his problem is to determine whether a newly written review that has not been voted by the users is potentially helpful or not. Our team analyzed the dataset provided by Dr Kevin, built an identifier by a 5NN machine learning algorithm, ran several tests about the identifier and designed a website extension prototype.

This project report will indicate an introduction and motivation for solving the problem, a problem definition explains the obstacles and corresponding solution, objectives that the project will achieve, a review of literature and related work that identify and describe briefly the results and any published work related to the report, a methodology part explains the team's methodology for solving the problem, a system architecture part introduce how the solution is structured, a software used for the system part, a prototype development shows the output, system documentation, issues arising during prototype development and prototype testing,

Furthermore, this project report will describe a test management that explains test plan for prototype, scope for testing, environment involves stakeholders, test procedures, non-functional testing and test result, a project recommendation indicating limitations and further development of the solution and a summary.

2. Introduction and Motivation

Review, a single feedback which is confided from consumer to their consumption or the specific good on the bill. In this fast-consumption era, the value of the feedback is not limited to businesses, also consumers cherish the comments. Instead of reading feedback from other users on the official site of the product, review sites vastly take advantage of the market to attract and content large range of consumers online. The diversity of the online reviews starts to confuse or even disturb consumers to making desirable decisions, then useful review become meaningful for modern consumers. Certainly, a useful review is the information from previous consumption that could inspire satisfaction of consumers while making a purchase decision that best fit their needs [7].

We’re more likely to make a purchase if others around us—even total strangers—agree that it is a good decision. Today, online reviews are the biggest source of social proof and have a significant impact on sales [9]. There is a problem arising around the project and pending for a resolution due to the derivative challenge from the traditional voting system on review sites. The existing system adopt rate on the reflections and suggest helpful review with the majority vote, otherwise, there is a shortage on recognizing newly review whether it is helpful. The society yearns an effective resolution to replace the voting system since there exist high dependency on the modern online review system. How important is customer review to shoppers? It turns out that this is very important. The fact is, 90% of consumers read online reviews before visiting a business. And 88% of consumers trust online reviews as much as personal recommendations [2].

Helpful review identifier is the project for furnishing suitable solution to both consumers and review sites online. There is finalized prototype to achieve the requirements form societies and provide appropriate integration of the new function and the sites. To be more attentive and satisfactory, the project constructs the identifier with appropriate methodology, tools and techniques to classify newly written review whether it is helpful for consumer and highlight the important portion from the original review to clarify the helpful attributes on the review.

3. Problem Definition

There are four questions related to the project:

- **How the newly review is helpful without voting, what method should we use to solve the question by our client?**
  To guarantee the newly review to be identified as unbiased helpful, a classifier is suitable to replace the inefficient voting system. Since we are provided a large number of data from the dataset provided by our client, a machine learning algorithm is suitable that machine learning algorithm can learn from past reviews, by specific attribute and after training process it can predict the new review and determine whether it’s helpful or not.

- **How do we analyze the dataset provided by our client, what attribute do we need to use for the project and why?**
  The classifier can learn from the previous reviews to deduce the latest comment is helpful or not individually, and the accuracy of the prediction depends on the reliability of well- structured data sets. According to the study of helpful reviews, those analysis reported review depth and extremity are massively taking influence on the review helpfulness. With the qualitative research from authority institution, our project group also processed quantitative analysis by collecting realistic statistics from 30 online respondents, the results from the survey state that, based on general perspectives, moderate comment length and rating scores, and commendatory terms are the most valuable features to evaluate a helpful review online. Thus, the precision of the core of the datasets are the specialized attributes mentioned above, where could stimulate accurate prediction for the newly reviews.

- **What should the user interface look like to convenient customer?**
  Since the project is asked to build a helpful review identifier, a user interface is also needed to present the results of classifier. Should the user interface be an application? a website? Or a plug-in browser extension? Since the production of project will interact with both review website and customer, the type of interface can affect the usability and how user-friendly to the customer. Functional browser extension integrate with the classifier can convenience readers to receive results from the current webpage, it is enabled to remove time cost from copying the reviews or transition between web pages.

- **How to facilitate decision making of review readers?**

  Since the consumers frequently search online reflections for some purchase decisions, to support instant decision, identify the useful feedback with a classifier consists of machine learning algorithms is promoting reliable quick response to readers.

4. Objectives

Intuitively, the identifying helpful review project's goal is to develop a real-time helpful review identifier to replace the voting system and facilitate consumer decision simultaneously. To foresee the future helpful review identifier, the objectives are declared in the following statements:

❖ To implement an identifier prototype with concurrent identifying function to determine online helpful reviews through IT techniques.

❖ To apply machine learning algorithm on the identifier in the design stage to fulfil the functionality assigned from original system and consumers, apply provided dataset to the algorithm and test its accuracy.

❖ To convenient consumers to identify useful comments on online review sites by building browser plug-in architecture amid design and implementation processes, since the project only has two month to achieve, a plug-in browser extension prototype is expected to design

❖ To establish a prototype at the end of the project being independent to the voting system, and bridging connection with legacy system to perform integration for problem resolution.

❖ To initiate project prototype development by preprocessing essential data resources from reliable sources, also guarantee the representative of the dataset after completely structured.

❖ To increase consumer satisfaction on decision making with conceptual identifier by managing user involvement in an acceptable scale and range of consumer types within the design and testing stage.

5. Review of Literature and Related Works

Before the implementation of the system, the qualitative analysis of similar review-identifier system, and finding some literature and related scholar articles were quite necessary steps, since these related references could actually help us generate some insights for prototype design.

There are two cases which help us to design the prototype, the first one is a website called Fakespot (http://www.fakespot.com) that aims to identify the reviews as well which is very aligned with the objectives to our project, and thus gave us some ideas about how to present the results of the identifier. The second case that recommended by our tutor Mr. Jindal is a chrome extension application called Grammarly, it provided us with penetrating practice on how to design the logo and the interface of the chrome extension, even our main functionality like highlighting the helpful sentence were inspired by it.

For the machine learning algorithm design part, there are three references that actually used in the algorithm design. First of all, it helps us to decide which machine learning algorithm will be used in the classifier. Based on the lecture materials of Introduction to Artificial Intelligence[4], it states that K-NN is one of the machine learning algorithms for predictive modelling problems, which is very suitable in this case since the aim of the classifier is to deal with dynamic data. The next step was, selecting the essential attribute of our model. According to the study of customer reviews on Amazon.com [8], they used five variables during their descriptive statistics analysis and another study of meta-review[9] claims that there are two essential review characteristics which are review length and review readability. Thus, in our algorithm, there are three attributes used for training the dataset which are the length of review text, star rate beside the review, structure of commentary text elements for the review readability, which is the number of nouns & adjectives words [3][11].

6. Methodology

Waterfall model is the methodology we choose during the whole project. The waterfall model describes a development method that is linear and sequential, it aims to complete the current activity before starting the next activity. Our project aims to build an extension for the users to identify a review is helpful or not, which have clear and fixed requirements. Our group understood the technology of machine learning and building the browser plugin. since the schedule for this project is around 10 weeks, waterfall is the most suitable choice.

Using waterfall model provide some advantages in the project. Due to the waterfall model is simple and easy to understand, use and manage, each phase has specific deliverables and a review process. it aims to complete the current activity before starting the next activity. Phases are processed and completed one at a time, they do not overlap. By using gantt chart, a schedule is set with deadlines for each stage of development and keep product be delivered on time.

There were 6 phases in the waterfall model during our project. Requirements gathering and analysis, system design, implementation, testing, deployment, maintenance.

7. Architecture

In the entire project development, there segregated the identifier into 2 proportions, which were data analysis model with unsupervised learning algorithm and a browser extension was to achieve the user requirement by basic prototype. Broadly, the identifier resembles 2-tier Client/Server Architecture to build up the connection between the potential online consumers and the project solution, that the Yelp database, the first-tier shared information with the identifier content, and another tier was approaching to client site. This separable application and client servers prevented the interruption of the user environment collapsed with the analyzing and testing environment. On the two components of the identifier, 5-Nearest Neighbors algorithm with Euclidean Distance Function [4][6] implemented analysis on the attributes, and then released relevant prediction on the clusters of the reviews, also the browser extension was designed to construct a comprehensive interface to allow online consumers to notice the helpfulness of a review and support their further consumption decisions.

7.1. Data Analysis Model

This data analysis model consisted of pre-processed data of reviews on Yelp, an unsupervised learning algorithm to cluster the similar features of online comments, and Euclidean Distance calculate the distances between two review elements, that simply extend to the *Single Link* function [6], which can make a prediction of whether the new review is helpful or not, based on the smallest pairwise distance between items from each cluster to define the corresponding similarity of the helpfulness. More dimensionally, each individual review would be distributed in a high-dimension area, that each attribute of the review represent a specific coordinate of the item, and the unique coordination of a comment would be defined by its values. Each voted review owned a known class, if a new unknown review joined in the area, we could identify its class is helpful or non-helpful by analysing the distances among the existing clusters, then determining 5 of the reviews that had the shortest distance to the new review, finally decided the class of the unknown comment by calculating the major class of the five nearest reviews.

7.2. Browser Extension

The browser extension was designed to contain viewable application to display the result of analysis on a web browser, that the interface could directly transfer the predicted results without voting, subsequently, the online consumer would be able to view the words we listed for them to interpret how the helpfulness was embedded in the current review texts, the amount of the important words would be scalable in order to make fast decisions based on the illustration from the interface.

8. Software Used for Project

During our project development period, we need various software to help us implement the project. Excel is the first software we used. Since we got the Yelps dataset include millions of records, we have to do the data processing and data mining to ensure we have the correct attributes that meet the criteria in our algorithm. Weka is also a useful software that can help us analyze the data and it is a collection of machine learning algorithms for data mining tasks. In other words, It contains functionalities for data pre-processing, classification, regression, clustering, association rules, and visualization. Therefore, we predicted nominal and numeric quantities using classifiers in Weka and training set is to evaluate based on how well it can predict the class of the instances it was trained on. We then split training set and testing set by 85% , 15% respectively. The training set, percentage split, supplied test set and classes are used for clustering, for which the user can ignore some attributes from the data set, based on the requirements. Then we found the correlation between attributes and value, we applied our testing set to find the accuracy of our 5NN algorithm. After building our prototype, we used google form to collect satisfaction feedback from our users.

9. Prototype Development

The processes of prototype development were initially committed by cleaning the raw review data from a large CSV file of Yelp textual reviews database, then determine the numeric attributes with some formulations and calculations in Excel and python coding. Before transferring the structured data into analysis model, that the data was certainly normalized in Weka in order to apply a general scale, without losing information about the helpfulness of the reviews or misrepresenting differences in the ranges of values. Inside the content of the data analysis model, the application of 5NN algorithm and Euclidean Distance were combined by programming logics in python, they were wisely learning the training set of the review and generated reasonable prediction to identify testing data as a newly written review is helpful to consumers or not, since the helpful features of the involved dataset was conducted from reliable research and investigation.

9.1. Required Output

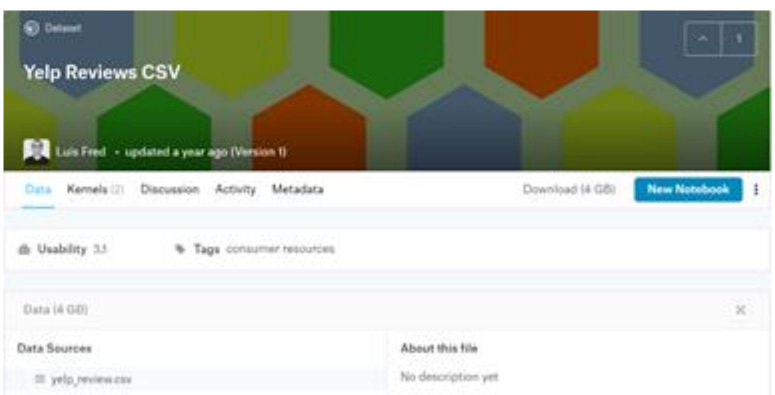
Prototype	**Required Resources**	**Exact Required Output**
Data Analysis Model	Review data from Yelp dataset	1. Raw data needed to extract from extremely large Yelp dataset. 2. Randomly selected enough reviews from CSV data file. 3. Data cleansing on unnecessary data that only “star”, “review” and “useful” remain for pre-processing. 4. Used text mining to process the “review” column, then obtain length and sum of adjective words and non-repeated nouns of each review. 5. The sum of the list of the specific words were defined as the value of the commentary texts. 6. Normalized the numeric “star”, “length” and “value” to acquire proper distribution for distance calculation.

	Machine learning algorithm to perform classification	1. Used Euclidean distance to measure the distance of each review with the three helpful attributes: “star”, “length” and “value”. 2. Use 5NN algorithm to isolate the reviews into two classes by firstly storing the distances of the training data, that is whether helpful or not helpful by grouping the similar values of the distances. 3. The first review forms a cluster of itself. 4. The following reviews were either merged with and existing group or form a new cluster of itself. 5. Each attribute contributed the same weight. 6. If distance (new review, existing review) < threshold then merge, else form a new cluster.
	Successful prediction of helpful review	1. Testing sets joined into the normalization. 2. Have the same distribution as the training sets. 3. The identifier was trained by 5NN algorithm, that the newly inserted testing review has no voting to support its helpfulness would be identified to a certain cluster depending of the closed five training reviews. 4. The python code of 5NN algorithm automatically report the diagnosis in an output file.
Browser Extension	User-friendly interface	1. Clean UI design 2. Easy recognized icon (magnifier and bulb icon) 3. Users should understand how extension works at the first time

	Identified result of certain reviews	1. Show the result classification of certain reviews (marked helpful or non-helpful), review that can not be recognized by the algorithm will show nothing . 2. Helpful words and sentences marked green
	Justification of helpful reviews being identified	1. By clicking the bulb icon, some helpful words shown to help user understand why certain review is identified as helpful. 2. Furthermore, by clicking the exclamation mark icon, more specific justification like length, count of helpful words will be shown.

9.2. System Documentation

ü Acquired trustful Yelp review dataset in CSV file format online.



ü Randomly selected 2000 reviews from the CSV file by using excel data model connection.

ü Only loaded the data with defined attributes: “star”, “review” and “useful”.

ü Applied Excel mathematical functions to calculate the length of every single review that transferred into the new CSV file.

ü Imported Natural Language Toolkit in python to retrieve non-repeated nouns and adjective words from the new CSV file created above.

```
import csv
import nltk
import sys

from collections import defaultdict

columns = defaultdict(list) # each value in each column is appended to a list

# function to test if something is a noun
is_noun = lambda pos: pos[:2] == 'NN'
# function to test if something is a adjective
is_adj = lambda pos: pos[:2] == 'JJ'
# do the nlp stuff
```

ü Count the value (sum of the number of non-repeated nouns and adjective words) of each plain review selected from the dataset.

```python
with open('review.csv') as f:
    reader = csv.reader(f)
    first_row = next(reader)
    for row in reader:
        list.append(row[1])
        tokenized = nltk.word_tokenize(row[1])
        nouns = [word for (word, pos) in nltk.pos_tag(tokenized) if is_noun(pos)]
        adj = [word for (word, pos) in nltk.pos_tag(tokenized) if is_adj(pos)]
        nouns_list.append(set(nouns))
        adj_list.append(adj)
        n = len(set(nouns))
        a = len(adj)
        total = n + a
        # print(total)
        rev.append(total)
        for (i,v) in enumerate(row):
            columns[i].append(v)


f = open("value.csv", "w")
```

ü Statistically measure the mean and median of the votes of useful column, that were both close to 1.5 votes on identified a useful comment.

ü Simply judged a voted review is helpful based on 2 polls on it, less than 2 was not being considered as helpful since there was a lack of attention on the review itself, that represented the feedback was meaningless to the public.

ü Restored all structured data in a formal CSV file named as “prepro.csv” that contained numeric values of star rating score, review length, review value, and a class helpful is determined whether yes or no.

ü Opened “prepro.csv” file in Weka, and ticked all the attributes to normalize, and then saved as “prepro_normalised.csv”.

ü Split the entire “prepro_normalised.csv” file into desired portion of training and testing sets and compile the “Identifier.py” to process training and testing simultaneously.

identifier.py | training.csv | testing.csv

```python

# predict each one new example vs all training data using 5NN
def five_NN(test, traindata):
    distance = []
    for trainsample in traindata:
        distance.append((euclideanDistance(test, trainsample),trainsample.split(',')[-1]))
    distance = sorted(distance, key=lambda x: x[0])
    #print(distance)
    # Extract the first k values. compare the number of yes and no, if yes>=no return yes, otherwise no
    all_result = []
    #print(all_result)
    for all_class in distance:
        all_result.append(all_class[-1])
    #print(all_result)
    #calculate the number of yes and no
    number_of_yes = 0
    number_of_no = 0
    for all_classes in all_result[:k]:
        if all_classes == 'yes':
            number_of_yes = number_of_yes+1
        if all_classes == 'no':
            number_of_no = number_of_no+1
    #print(number_of_yes)
    #print(number_of_no)
    if number_of_yes >= number_of_no:
        f.write('yes\n')
    else:
        f.write('no\n')
```

9.3. Issues Arising

➢ **Uneven number of helpful and non-helpful dataset**

Within the development of the data analysis model, the functionality of the machine learning algorithm was dubious, consumers could take an unconscious guess on the result of the classification, that the training set includes nearly 70% reflections were non-helpful reviews and the rest were helpful, where the accuracy on the current dataset was not representative due to low chance on the estimate of helpful review. Therefore, agreements were contracted with the client, on the number of two opposite classes, 1000 of helpful review and 1000 of non-helpful reviews were randomly extracted from the Yelp database, that promoted an even chance for forecasting the helpfulness of the newly updated review. Lastly, refined arrangement of the dataset enhanced the usability of the data analysis model, that emphasized the 5NN algorithm could provide prediction instead of random guess based on the portion of each class of the reviews.

➢ **UI design (icon implement)**

At the beginning stage, we used red exclamation mark at the end of useful reviews. Then we found we should maintain consistency between marked words and sign, as a result, we change the red exclamation mark to green exclamation mark to match the green highlighted words.

Next step, we found that we need to give user straight sense whether the review is helpful or non-helpful. We added a seal with red word ‘approve’ in it. Then for the same reason (match the green mark), we change the red seal to green seal with word ‘helpful’ and change the green exclamation mark to a green bulb icon in order to build a more user-friendly design.

- **Function shown in the interface**

First, we only marked some keywords green in the helpful reviews and some keywords red in the non-helpful reviews. However, during some users feedback, they have no idea what exactly these words mean. So we changed to marked all content in the helpful reviews green and all content in the non-helpful reviews red. But this still not satisfied our client needs, he said red will get more attention to user than green, and it will make user confused. As a result, we deleted the red mark in non-helpful review and focus on helpful reviews. Furthermore, users do not understand what is the keywords mean in the extension. We changed the ‘keywords’ into ‘helpful keywords’ and give the specific explanation of these words for users to understand.

9.4. Modified Prototype

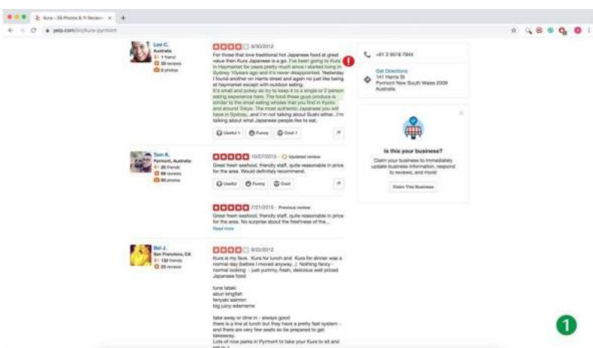
Original Prototype	Modified Prototype

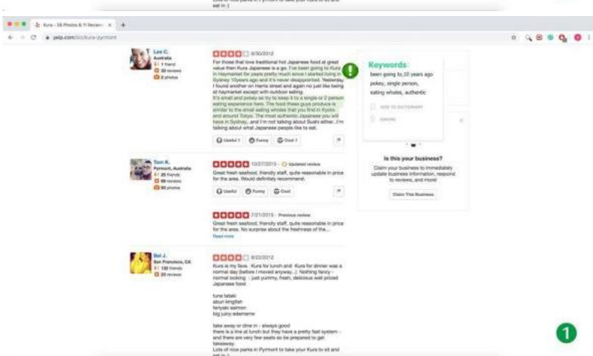
	stars	length	value	helpful				
2	0.75	0.547619	0.27907	no		yes	0	0.6175
3	1	0.3	0.155039	yes		no	0	
4	1	0.309524	0.162791	no		no	1	
5	0.5	0.52381	0.170543	yes		no	0	
6	1	0.519048	0.271318	no		no	1	
7	0.75	0.590476	0.379845	no		no	1	
8	0.75	0.52381	0.24031	yes		no	0	
9	0.5	0.757143	0.317829	no		no	1	
0	0.5	0.6	0.302326	no		yes	0	
1	0.5	0.895238	0.372093	yes		no	0	
2	0.5	0.438095	0.224806	yes		yes	1	
3	1	0.409524	0.224806	no		no	1	
4	1	0.461905	0.209302	no		no	1	
5	0.75	0.638095	0.263566	yes		no	0	
6	0.75	0.833333	0.434109	yes		no	0	
7	1	0.704762	0.379845	yes		no	0	
8	1	0.352381	0.24031	no		no	1	
9	0.25	0.519048	0.193798	no		no	1	
20	1	0.204762	0.085271	no		no	1	
21	1	0.142857	0.116279	no		no	1	
22	0.75	0.714286	0.379845	yes		no	0	

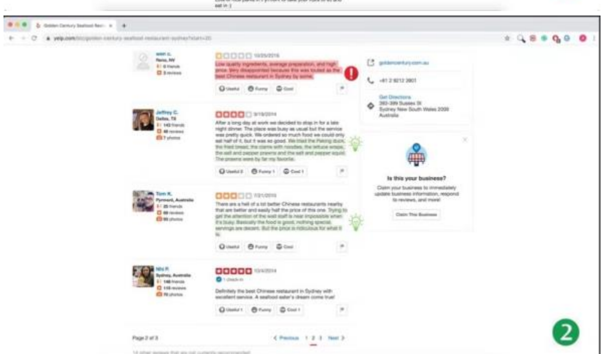
425	1	0.661905	0.348837	no				
426	0.75	0.185714	0.131783	no		no	1	0.636
427	0.75	0.371429	0.170543	no		yes	0	
428	1	0.380952	0.162791	no		no	1	
429	0.75	0.057143	0.046512	no		no	1	
430	0.75	0.080952	0.046512	no		no	1	
431	1	0.219048	0.155039	no		yes	0	
432	0.5	0.295238	0.124031	no		no	1	
433	1	0.071429	0.046512	no		yes	0	
434	1	0.585714	0.310078	no		no	1	
435	0.75	0.209524	0.124031	no		no	1	
436	0.5	0.147619	0.069767	no		no	1	
437	0.75	0.047619	0.03876	no		no	1	
438	0.75	0.390476	0.162791	no		no	1	
439	0.5	0.12381	0.046512	no		no	1	
440	0.75	0.042857	0.031008	no		yes	0	
441	1	0.380952	0.186047	no		no	1	
442	0.75	0.095238	0.069767	no		no	1	
443	0.75	0.090476	0.031008	no		no	1	
444	0.25	0.247619	0.139535	no		yes	0	
445	1	0.290476	0.147287	no		no	1	
446	0.5	0.1	0.054264	no		no	1	

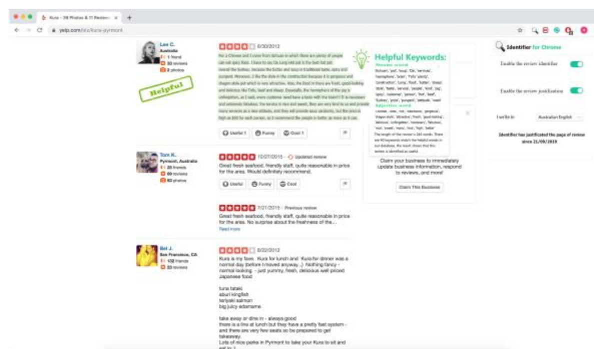
The previous prototype of included approximate 70% non-helpful reviews, and roundly 30% helpful reviews, the estimation is lower than a general prediction based on the dataset itself.

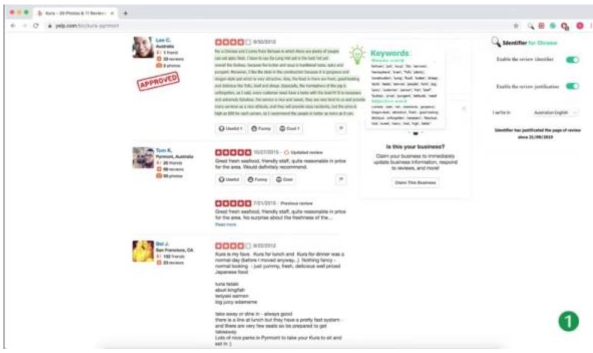
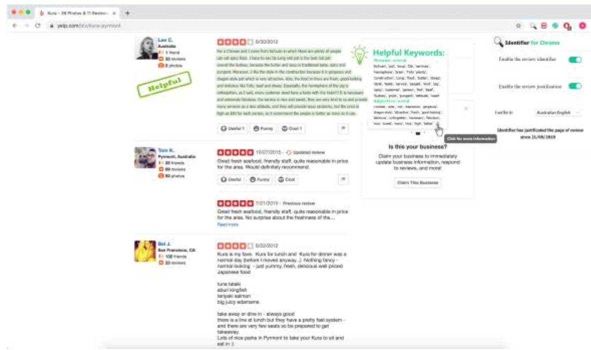
The rearranged prototype combined 50% non-helpful reviews and 50% helpful reviews, the effects of the classification from the identifier is obviously increased, that the prediction accuracy is higher than a guess on data portions.









The previous prototype using different icons to classified helpful and non-helpful reviews, however some users still confused and mislead by what exact these icons mean.	The modified prototype using seal to show users the review is identified helpful or non-helpful. Green bulb icon makes user more likely to check the justification.
	
The previous prototype using red 'approval' seal which make user puzzled the exact meaning of that. The color does not match the green marked words.	The modified prototype using green 'helpful' seal. By clicking the more information button, user can check more specific explanation of these helpful keywords.

9.5. Prototype Confirmation Requirement

Fulfilment of the prototype development had decomposed the problem resolution in this project and drove the progression into a testing stage. To successfully deliver the IT solution to client, there settled several confirmation requirements to achieve the request from planning and early design stages:

- The attributes are correlated to the fixed classes.
- The performance of the data analysis model of the identifier should be within or above the benchmarks of the classifiers.
- The accuracy should be maintained when a real written review is added into the identifier.
- The prognosis should be released and displayed for online users.
- User interface is approaching to assist real online consumers decision-making.
- Helpfulness is identified and should be understood by common online consumers.

10. Test Management

10.1. Test Plan

The aim of the test is to develop the prototype iteratively through the feedback of user and algorithm. Basically, test plan has been divided into 5 stages shown below:

10.1.1 Analyze the product

In the first place, the product under test is the extension of helpful review identifier, since it is necessary for us to research the demand and expectation of clients and market from this application. Then, we have decided the approach to analyze the product in a circulation below:

1. Interview client, designer and developer
2. Review the product and project documentation
3. Perform product walkthrough

After that, there are several questions appeared:

- Who will use it?
- What is it used for?
- How will it work?

After the completion of research through the survey in the form of questionnaire posted online, it has been identified the audiences are mostly those people who are used to checking the review for making the decision in daily life. The extension is used for identifying the helpful review and highlight the paragraph in the review-based website which is mostly Yelp. The manipulation of the application has been mostly introduced before in which extension use the machine learning algorithm to identify the word review.

10.1.2. Scope of Testing

In Scope:

- Build up the most expected construction and design of application through testing
- Test the user interface in prototype of the script in terms of learnability and consistency
- Test the non-functional characteristics of prototype
- Usability testing in terms of effectiveness, efficiency, memorability
- Test the feasibility and reliability of the prototype of plugin browser extension
- Collect the satisfaction of stakeholders who are client, user and developer
- Train and test the feasibility of machine learning algorithm through the Yelp dataset
- Test the accuracy of machine learning algorithm

Out of Scope:

- Test the reliability of the dataset

- Test the accurate number of budget
- Identify and test the category of risk in the project
- Test the running speed of the algorithm
- Test the stability of running the machine learning algorithm

10.1.3 Identify Testing Type

Our group decided to take integration testing, a level of software testing, test the individual units integrated as a group. This level of this testing is supposed to explore the faults in the interaction between combined units which are referred to the machine learning algorithm and the user interface of extension.

10.1.4 Define Test Objective

To define test objectives, The first stage is to list the whole features of application that need to be tested which is functionality, usability, user interface.

Functionality:

- Ability of self-learning by machine through the identification of attribute
- The stable accuracy of the calculation under various dataset
- The most functional algorithm of 5NN compared to 1NN, 3NN
- The prediction of the helpful or non-helpful review
- Display the result of helpful or not through the running of algorithm

Usability:

- User friendly control and understand
- Efficiency, memorability, learnability and consistency

User Interface:

- Proper plugin element as button, menu, text and image
- Clear contents with the justification of the result
- Plugin layout design

Based on those features has been listed above, we can define the test objective of the project helpful review identifier as the following:

1. Check and update the functionality can mostly meet without any error and improve the accuracy iteratively to meet 70%
2. Verify the usability of the extension, prove this functionalities are useful and convenient to user
3. Test the Plugin browser extension meet the demand of stakeholders and make sure the running of user interface is expected

10.1.5 Define Test Criteria

The choice of test criteria is suspension criteria which elaborates the critical for a test. the circulation of active test will be suspended until the criteria are resolved if the target of testing meet the suspension criteria during testing[1]. Basically, this is a criteria to check and fix the certain testcase iteratively until the error has been fixed in the aspect of algorithm or it meets the successful criteria which is the satisfaction of clients under the prototype of interface. In other words, the test criteria in the project is iterative update and improve in terms of prototype.

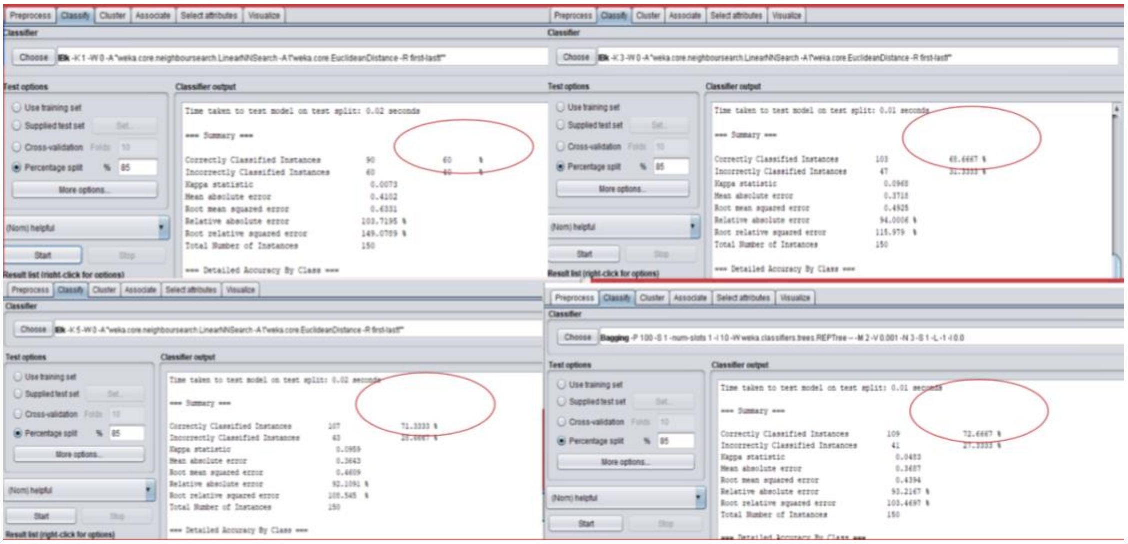
10.2. Stakeholders

Stakeholders are referred to people who have direct or indirect interest in the project, then we can deduce three essential roles of stakeholders who are client, user and developer. Accordingly[10], involvement of stakeholders has played a crucial role in testing, because people can tell what they prefer and what they disliked to display on the prototype, that is, stakeholders provide their feedback and the organization collect this feedback and do the modification to close the need of stakeholders. Client has the enormous embedded relationship and interest in which he wants to authorise and purchase the product, since they often offer their opinions and feedback which is important for us to update the extension iteratively. For example, client brought the question ‘How to user without any knowledge know how to use the application?’, then we used suspension criteria to make modification until we met the satisfaction of client. User is the customer and audience who is willing to use the application and their feedback is also crucial for us. Developer(end user) who work for the programming and developing the algorithm in the aspect of software systems and they need to update the application through the feedback from the user and client or their own testing.

10.3. Test Procedures

After the made of test plan, we have implemented the testing into the prototype which is machine learning algorithm and user interface, then the testing of algorithm plays a crucial role in test procedures. This part introduces the procedures of KNN testing, basic testing process, complex testing for reliability and testing stakeholders.

Firstly, we have tested the most feasible functionality of 5NN compared to 1NN, 3NN. Basically, we do the step before to analyse the testing set but we place the same data into different algorithm normalization sorted by weka separately, then we find the highest accuracy in 5NN and the lowest in 1NN which is 72.66% and 60% respectively. the accuracy of 3NN is 68.66%, thus the result demonstrates the most feasible algorithm is 5NN.



Secondly, this section illustrates the basic testing of the algorithm in detail. In the beginning, we extracted the data from Yelp dataset so it meant by those data have been identified as the helpful review or not. After preprocessing and normalising the dataset, we splitted this analysed data into two datasets which is training set and testing set separately. After modification through the feedback of client, the percentage of helpful and non-helpful is balanced as 50% on both sides because the output will be meaningless if they are unbalanced.The percentage of training set and testing set is respectively 85% and 15% of the normalized data because the training set needs plenty of data to support 5NN algorithm in the database, also this allocation has been proved as the high accuracy in testing[4]. We have already known the dataset is helpful or not from Yelp dataset, then the algorithm will calculate the average of helpful from 5 nearest cluster. Therefore, the algorithm will display the corresponding output with the percentage of helpful review of 5 nearest cluster. At last, we collect the results and compare to its original defined result in Yelp dataset then we can gain the accuracy.

Thirdly, for the reliability of testing, we also test the various datasets into the algorithm to avoid contingency in dataset, this paragraph illustrates complex testing in different dataset and the accuracy. Firstly, we tested 1000 number of review in total set, then we gain 63.6%, since the output is unexpected then we suspended it until gain the satisfaction of accuracy, then we tried inputting the dataset with the proportion of helpful review compared to non-helpful one is 1:2, and the total number is 2000, then we gain 65%, but we found the unbalanced helpful number is meaningless so we cross it out this solution. We tried identifying test it under balanced helpful number and get accuracy is 61.7% which is worse. Then we tried consider to improve the proportion in testing set to 50% compared to training set. After that, we gather 74% accuracy to meet the satisfied criteria.

At last, we need to test stakeholders, we also test the algorithm meet user's need in which we ask users to write their own review and give the solution of the review which is helpful or non-helpful, after we input this review as testing set and the training set is original 1000 Yelp dataset, the output of solution gain the accuracy is 50% but most of users think the identifer make sense because they don't believe their review is helpful exactly even they write it as helpful review.

10.4. Non-Functional Testing

10.4.1. User Interface

Basically, user interface we have demonstrated to user is screenshot containing the work of plugin browser extension in Yelp website and user feedback is the main way for us to test user interface.

The satisfaction of client is our successful criteria in this part then we bring the question to the user whether they are satisfied with the result of identification by extension in an online survey. Basically, we have done 2 testings in user feedback, the second testing is based on the suggestion from client. From the response, there is 72.2% rate representing satisfied and very satisfied in the total of 54 response and the rest of them are mark of okay and there is 76.5% rate of very satisfied for design of the extension. There is 74% users think the extension is useful to identify the helpful review. Therefore, We mostly gain the satisfaction of the prototype demonstration from user. However, client brought the question that how to let user know the high-lighted paragraph represent the meaning of helpful review without any knowledge, based on this suggestion, we add the icon with 'helpful' to support user to understand, and make the interface consistency with green colour which is a metaphor for working the extension.

Nevertheless, we decided to do another testing based on client's feedback which contain more questions in the online survey in which we ask user that can they know the meaning of high-lighted part and think the reason is easy to understand the reason. The second question is unexpected as 11.1% users think it is not easy to understand the reason. Then, we changed the content in display frame of the reason of helpful review.

10.4.2. Usability

In brief, we need to test the usability in terms of efficiency, memorability, learnability and consistency. Basically, we do another testing in the forms of online survey.

From the part of efficiency, we brought the question “Do you think the design of the extension can save the time?” and there is a rate of 88% which is very satisfied. 92% of users believe that it is easy for them to memorise the steps to use this extension which prove the memorability and learnability. Consistency played a crucial role in the usability because it has the most strong relevant with the interface of extension in which green color show up the interface main page and all the work of the prototype. Therefore, the question can prove the consistency is “Do you think the interface keep the same style” and “ Can you understand the meaning of green colour high-lighted paragraph without hint”. The solution is above 90% of users agree those questions.

10.4.3. Test Results

After finishing the testing, there are several results we have achieved:

- The highest accuracy with 74%
- Objectives confirmed by client
- The average of user satisfaction is above 80% through the survey
- Achievement of objectives in plan in terms of functionality, usability and user interface
- Success of implementing the algorithm as the prototype

11. Recommendations

11.1 LIMITATIONS

After meeting with our client Dr Kevin Kuan for several times and evaluated the project over and over again, there are still few limitations exists in this system.

For the part of prototype, the interface of the browser extension is only designed for displaying the results of the helpful reviews, therefore the user could only see the functionality when the helpful review appears. Hence in some situation, when they turn on the chrome extension, they might see nothing changes at the current page if there is no helpful review exists. So it might make the user a little bit confused with why there is no unhelpful indicator.

For the algorithm part, at the current stage, due to the limitation of the technique ability of the team. the algorithm only includes several attributes which are the number of nouns and adjectives, length of review and star rate beside the review. In other words, all the displayed results based on these three attributes, which might be a little bit inaccurate. Moreover, the data size used to train the algorithm is also a little bit small which is approximately two thousand reviews at the present stage. Overall, we still have a long way to improve our algorithm.

11.2 FUTURE DEVELOPMENT

Based on the limitation of the current project, there is some potential weakness of our project that could be developed in the future. First of all, the interface needs to be updated in order to enable results also displayed on the unhelpful reviews. In this case, all the non-helpful review text will be labelled as unhelpful in order to differentiate from the helpful reviews. In addition, the criteria for the results will also be shown in order to help the user understand the reason behind it and further trust the displayed result. Secondly, the algorithm also needs to be developed as well. In the future, more attributes like review readability or uploaded pics will be added to the algorithm to identify the helpfulness better. Thirdly, a really big dataset that includes at least half-million samples will also be used to train the machine learning algorithm, in order to improve the accuracy rate. Furthermore, based on the suggestion of Dr Kevin Kuan, the results will also be shown in four levels which are very helpful, medium helpful, less helpful and non-helpful.

Overall, after the development of both algorithm and prototype, we believe plenty of users will start using our application while browsing the review website. Based on this, the team of Group 10 could collect more realistic feedback and then evaluate the product and improve its performance in the long term.

12. Summary

As a result, Dr Kevin Kuan is basically satisfied with our algorithm and prototype since all functional and non-functional requirements have been achieved. The application delivers business value by helping users to identify whether the new review is helpful or not without voting system. Moreover, the platform (chrome extension) used to deliver the service also brings the uniqueness of our product and even more convenience to the user. More importantly, the core of the product is the algorithm design since it decides how the identifier works. And in order to meet Kevin’s expectation, the extension applies one of the machine learning algorithm called 5-Nearest Neighbors algorithm to train the dataset and thus it generated a reasonable prediction on the newly written review. After the algorithm and prototype design, the follow-up testing process also makes sure the performance of the system could meet the user's requirement.

Although the system is delivered at the expected time without any delay and basically meet Kevin's expectations, there are still some limitations and potential weakness on the usage. In the future, our team will keep evaluating the system by optimizing the algorithm and updating the prototype. Therefore, the service delivered by the system could be maintained at a high quality and thus attracting and helping more users to use our system while browsing the review website.

13. References

[1] How to Create a Test Plan. (2019). Retrieved from

https://www.guru99.com/what-everybody-ought-to-know-about-test-planing.html

[2] Khalid Saleh. *The Importance Of Online Customer Reviews [Infographic]* Retrieved from https://www.invespcro.com/blog/the-importance-of-online-customer-reviews-infographic/

[3] Keenan, E. (2009). Existential sentences in Tagalog: commentary on the paper by Joseph Sabbagh. *Natural Language & Linguistic Theory*, *27*(4), 721–735. https://doi.org/10.1007/s11049-009-9080-6

[4] Koprinska,I. (2019, April). Lecture 5, COMP3308 Artificial Intelligence. Lecture 5, COMP3308 Artificial Intelligence. Sydney.

[5] Koprinska,I. (2019, April). Lecture 6b, COMP3308 Artificial Intelligence. Lecture 6b, COMP3308 Artificial Intelligence. Sydney.

[6] Koprinska,I. (2019, April). Lecture 12, COMP3308 Artificial Intelligence. Lecture 12, COMP3308 Artificial Intelligence. Sydney.

[7] Miller, D. (1995). *Acknowledging consumption a review of new studies*. London: Routledge.

[8] Mudambi, S., & Schuff,D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com.(Case study). MIS Quarterly,34(1), 185–200.

[9] Smith Willas. *7 Reasons Why Online Reviews are Essential for Your Brand*. Retrieved from https://mention.com/blog/online-reviews/

[10] STC Admin. (2015). Involvement Of Stakeholders In Testing. Retrieved from http://www.softwaretestingclass.com/involvement-of-stakeholders-in-testing/

[11] [Review of *Review 2 -- No Title*]. (1877). *The Eclectic Magazine of Foreign Literature (1844-1898)*, *26*(3). Retrieved from http://search.proquest.com/docview/89758355/

[12] Wang, Y.,Wang, J., & Yao, T. (n.d.). What makes a helpful online review? A meta-analysis
of review characteristics. Electronic Commerce Research, 19(2), 257–284.
Retrieved from https://doi.org/10.1007/s10660-018-9310-2

14. Appendix

Survey to collect the requirement in the preprocess of the project

3. If the answer to the Q2 is Yes, How many words do you usually leave?

Mark only one cycle

◯ < 10
◯ 10 - 50
◯ 50 - 100
◯ >100

4. If the answer to the Q2 is Yes, Do you leave pictures as well?

Mark only one cycle

◯ Yes
◯ No

2. What is the real data size for this website or app to deal with?
Mark only one cycle

◯ <100
◯ 100-500
◯ 500-1000
◯ >1000

4. Will the user need to login to use the service?
Mark only one cycle

◯ Yes
◯ No

Online survey to test the user

User Survey

Dear respondent,

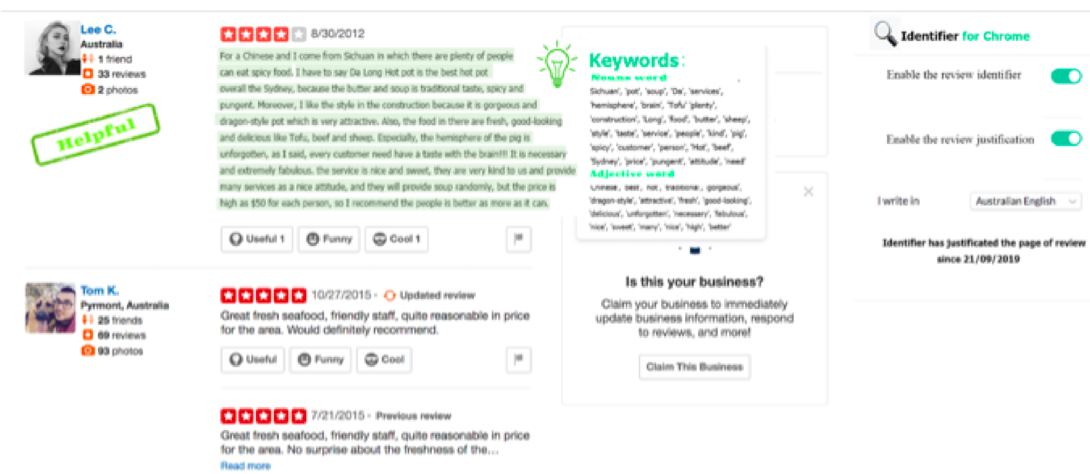
The aim of the questionnaire is mainly to collect the user feedback of algorithm and interface. Your response is very important to help us identify helpful reviews or non-helpful reviews.

We will not receive your name and address, and the credit grantor will not see your answers to this questionnaire. Your responses are anonymous and it will remain confidential.

The whole questionnaire will last no longer than 10 minutes, We would greatly appreciate your participation in this survey. Please help us by completing the questionnaire today.

Sincerely yours,
Group 10.

Image title



Are you satisfied with the results of identification given by the extension? *

	1	2	3	4	5	
Very well	○	○	○	○	○	Very bad

What is your judgment criterion of the last question?

Long answer text

Do you like the design of extension? *

	1	2	3	4	5	
Very well	○	○	○	○	○	Very bad

Do you think the extension is useful to identify the helpful review? *

	1	2	3	4	5	
Very helpful	○	○	○	○	○	Very unhelpful

Can you know the high-lighted review is meant by helpful review? *

- Yes
- No
- Maybe

Do you think the reason of the helpful review is easy understanding?

1 2 3 4 5

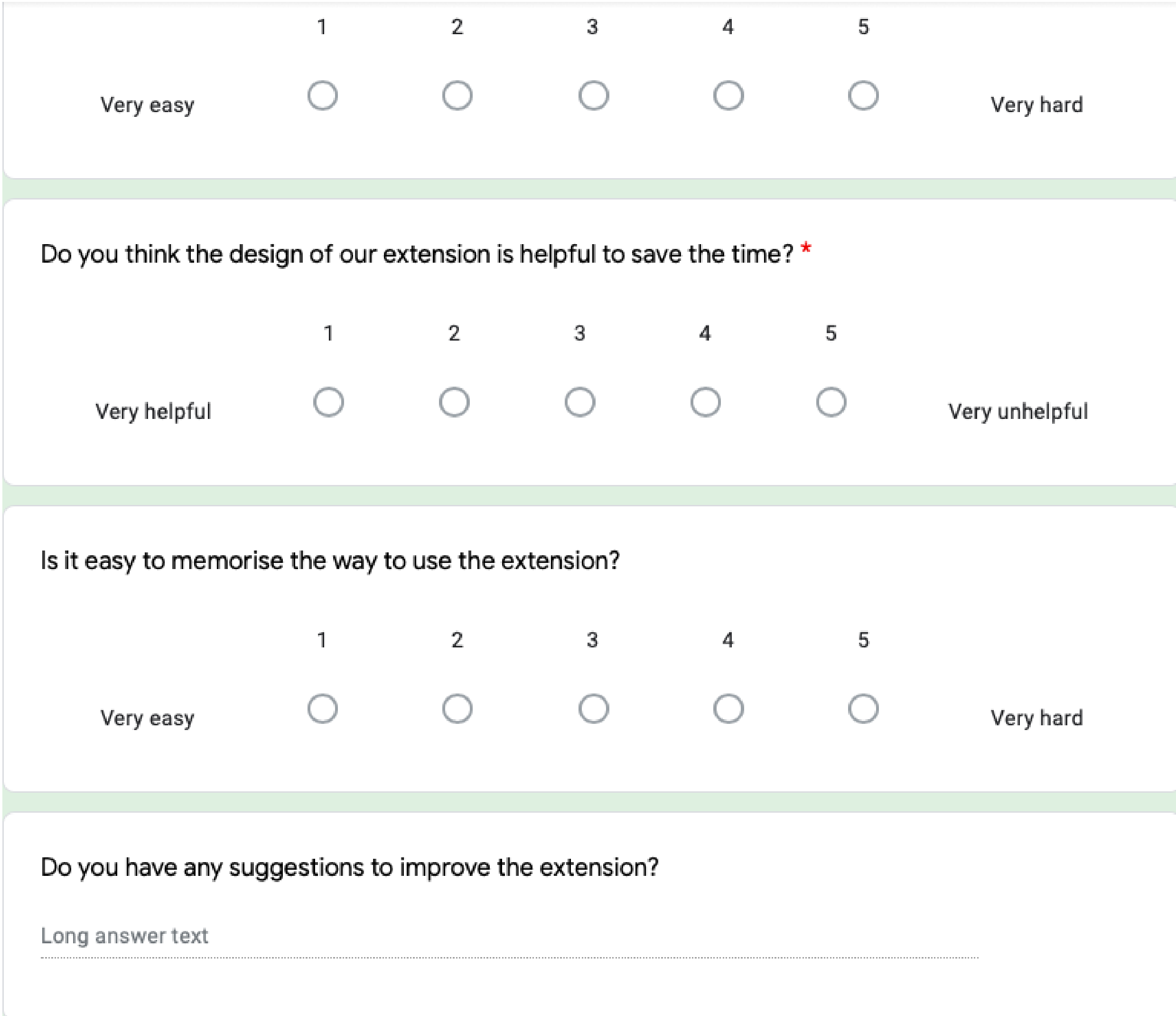
very well ○ ○ ○ ○ ○ very bad

Do you think the extension and the work keep the same style? *

1 2 3 4 5

Very helpful ○ ○ ○ ○ ○ Very unhelpful

Is it easy to know the icon of green bulb is the meaning of checking the justification? *



	1	2	3	4	5	
Very easy	○	○	○	○	○	Very hard

Do you think the design of our extension is helpful to save the time? *

	1	2	3	4	5	
Very helpful	○	○	○	○	○	Very unhelpful

Is it easy to memorise the way to use the extension?

	1	2	3	4	5	
Very easy	○	○	○	○	○	Very hard

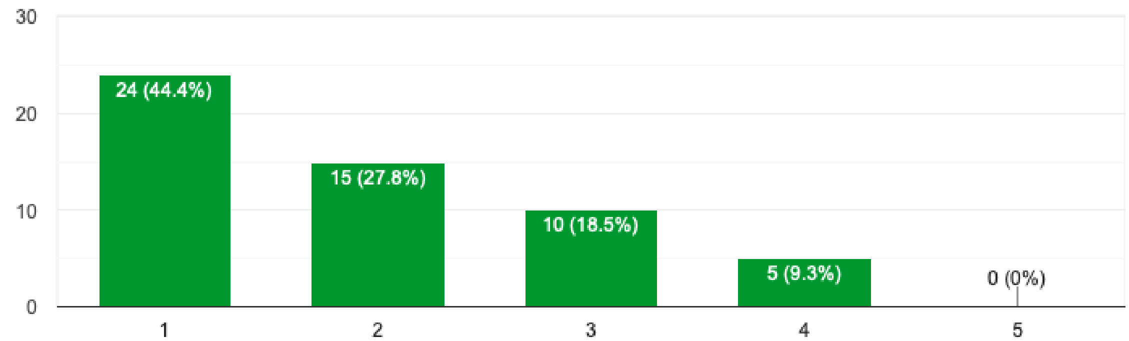
Do you have any suggestions to improve the extension?

Long answer text

The result of online survey

Are you satisfied with the results of identification given by the extension?

54 responses



What is your judgment criterion of the last question?

27 responses

USER-FRINENDLY INTERFACE

The number of star

it makes sense

The results meets my expectation

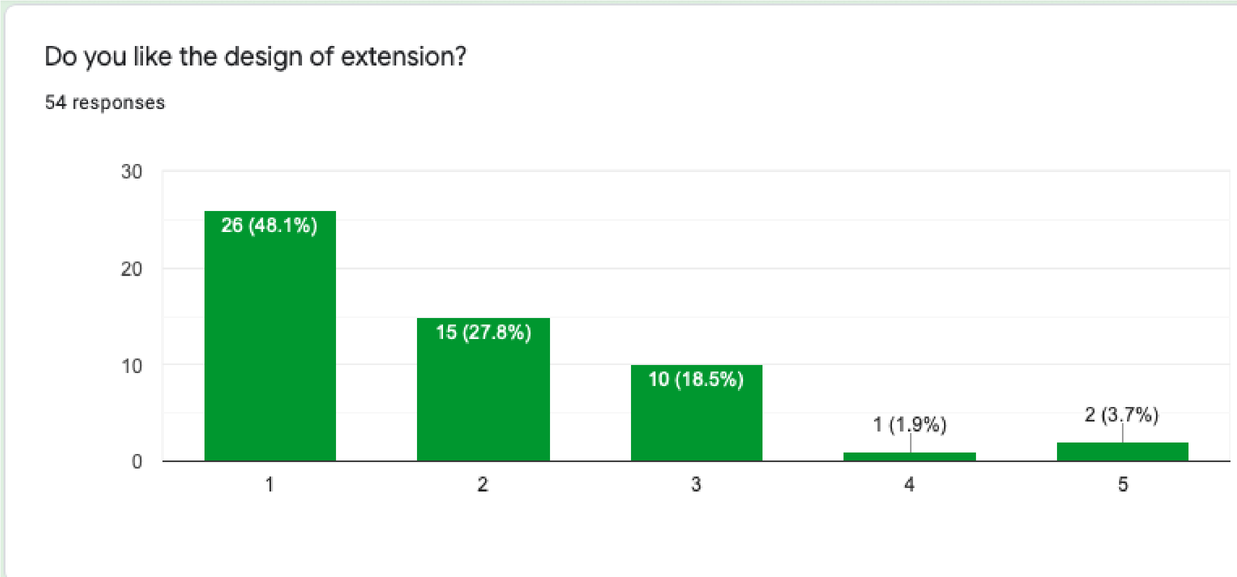
really useful to distinguish helpful and unhelpful reviews

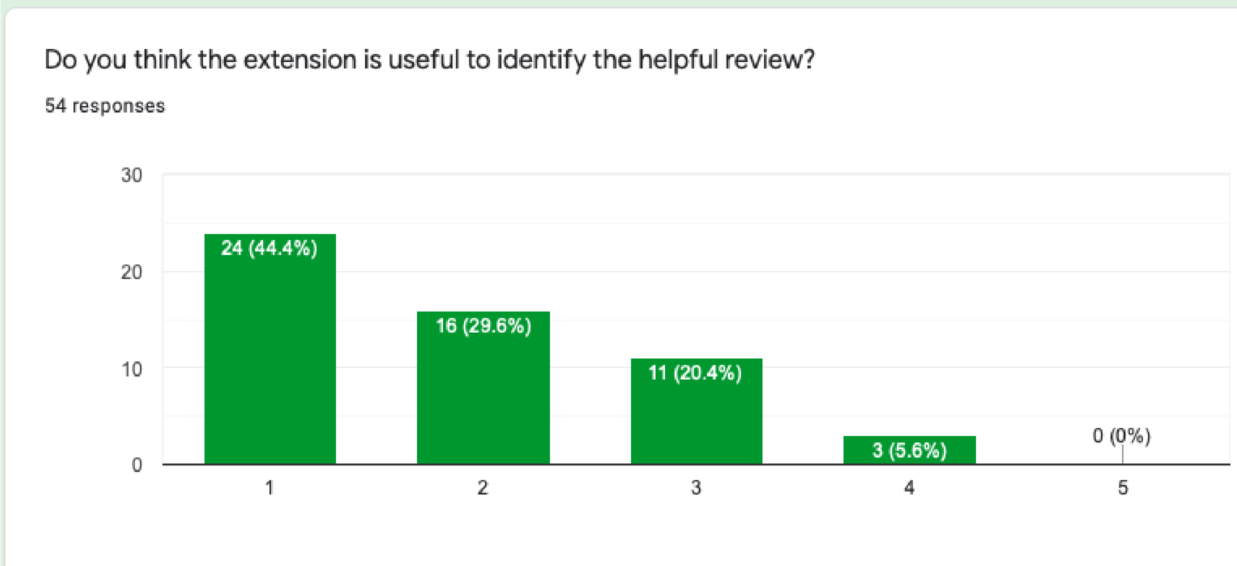
not bad

the prototype is clear

good

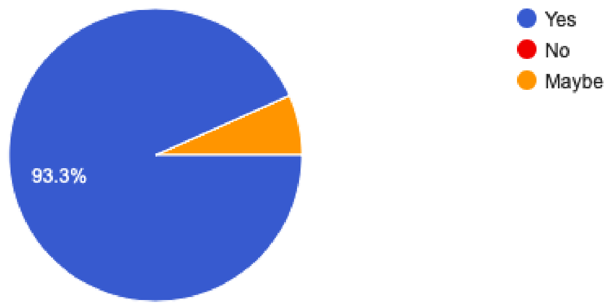
love the result, very helpful


Do you like the design of extension?
54 responses
30
20
10
0
26 (48.1%)
15 (27.8%)
10 (18.5%)
1 (1.9%)
2 (3.7%)
1
2
3
4
5


Do you think the extension is useful to identify the helpful review?
54 responses
30
20
10
0
24 (44.4%)
16 (29.6%)
11 (20.4%)
3 (5.6%)
0 (0%)
1
2
3
4
5

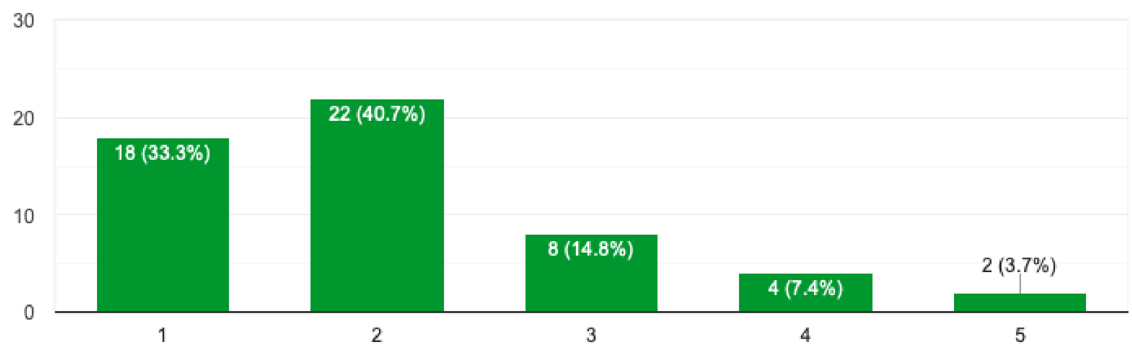
Can you know the high-lighted review is meant by helpful review?

30 responses



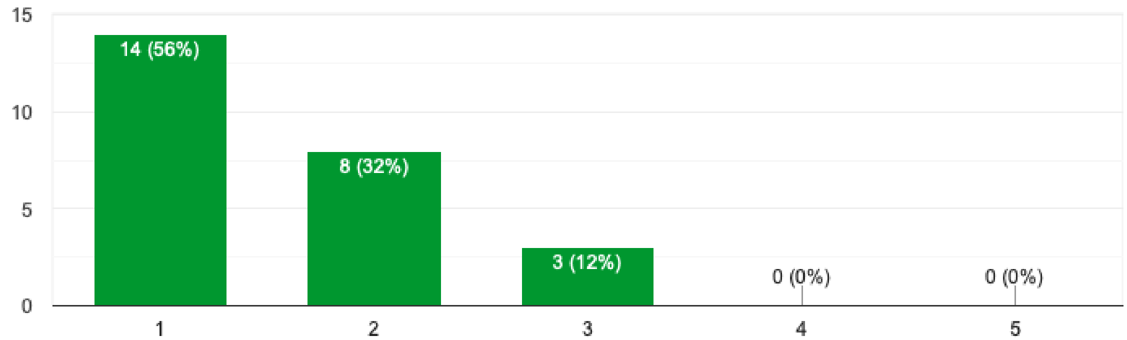
Do you think the reason of the helpful review is easy understanding?

54 responses



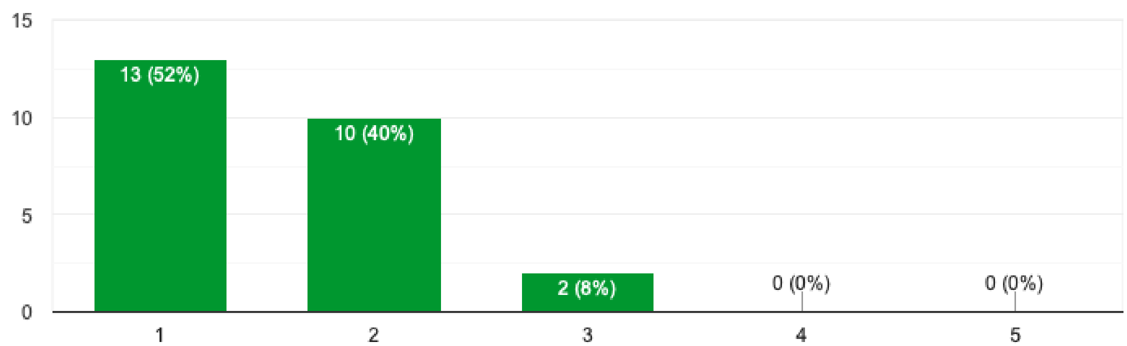
Do you think the design of our extension is helpful to save the time?

25 responses



Is it easy to memorise the way to use the extension?

25 responses



Do you have any suggestions to improve the extension?

21 responses

GOOD PROTOTYPE

The algorithm Need to be improved

Interface and algorithm works well together

the function helps me a bit, but not much

it can help me find the helpful reviews

very well designed, like it!

The green exclamation point is confusing

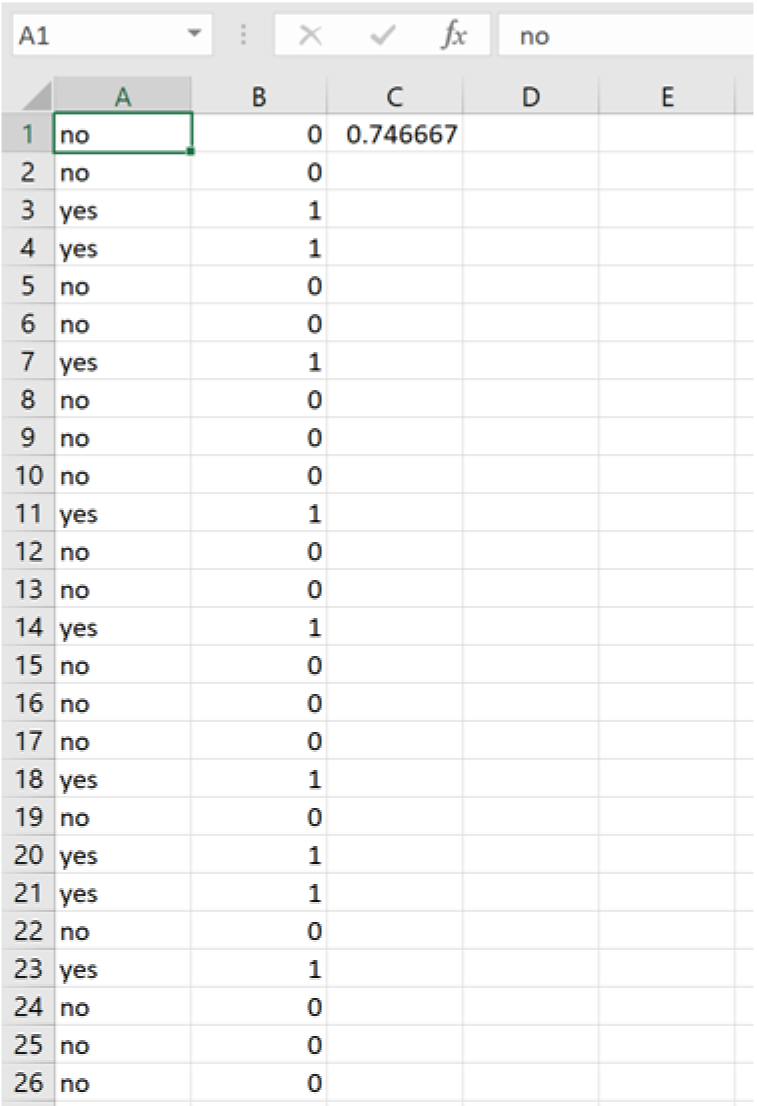
love it! Keep improving!

The algorithm

The accuracy through the testing set:

425	1	0.661905	0.348837	no				
426	0.75	0.185714	0.131783	no		no	1	0.636
427	0.75	0.371429	0.170543	no		yes	0	
428	1	0.380952	0.162791	no		no	1	
429	0.75	0.057143	0.046512	no		no	1	
430	0.75	0.080952	0.046512	no		no	1	
431	1	0.219048	0.155039	no		yes	0	
432	0.5	0.295238	0.124031	no		no	1	
433	1	0.071429	0.046512	no		yes	0	
434	1	0.585714	0.310078	no		no	1	
435	0.75	0.209524	0.124031	no		no	1	
436	0.5	0.147619	0.069767	no		no	1	
437	0.75	0.047619	0.03876	no		no	1	
438	0.75	0.390476	0.162791	no		no	1	
439	0.5	0.12381	0.046512	no		no	1	
440	0.75	0.042857	0.031008	no		yes	0	
441	1	0.380952	0.186047	no		no	1	
442	0.75	0.095238	0.069767	no		no	1	
443	0.75	0.090476	0.031008	no		no	1	
444	0.25	0.247619	0.139535	no		yes	0	
445	1	0.290476	0.147287	no		no	1	
446	0.5	0.1	0.054264	no		no	1	

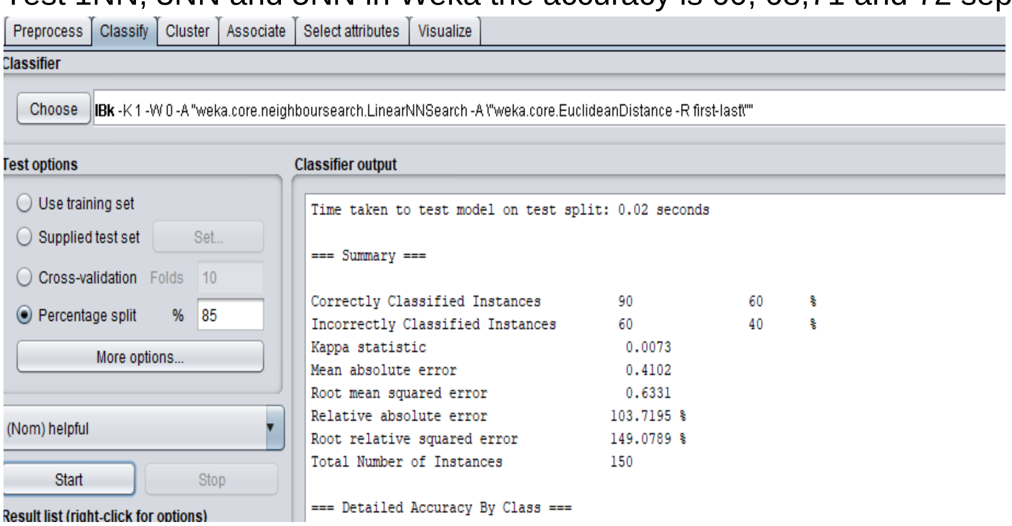
star	length	value	helpful						
0.75	0.545	0.261905	no				no	1	0.713333
1	0.285	0.134921	yes				no	0	
1	0.295	0.142857	no				no	1	
0.5	0.52	0.150794	yes				no	0	
1	0.515	0.253968	no				no	1	
0.75	0.59	0.365079	no				no	1	
0.75	0.52	0.222222	yes				no	0	
0.5	0.765	0.301587	no				no	1	
0.5	0.6	0.285714	no				no	1	
0.5	0.91	0.357143	yes				no	0	
0.5	0.43	0.206349	yes				no	0	
1	0.4	0.206349	no				no	1	
1	0.455	0.190476	no				no	1	
0.75	0.64	0.246032	yes				no	0	
0.75	0.845	0.420635	yes				no	0	
1	0.71	0.365079	yes				no	0	
1	0.34	0.222222	no				no	1	

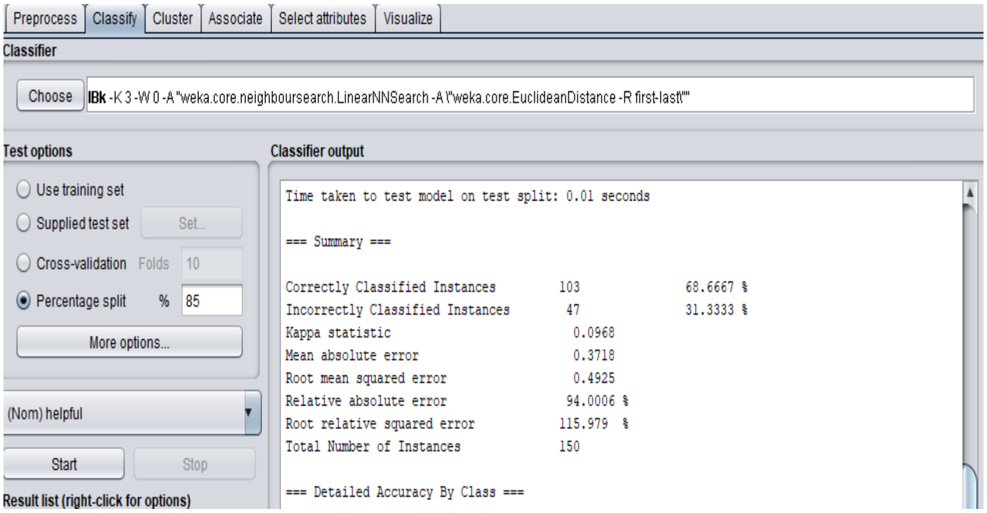


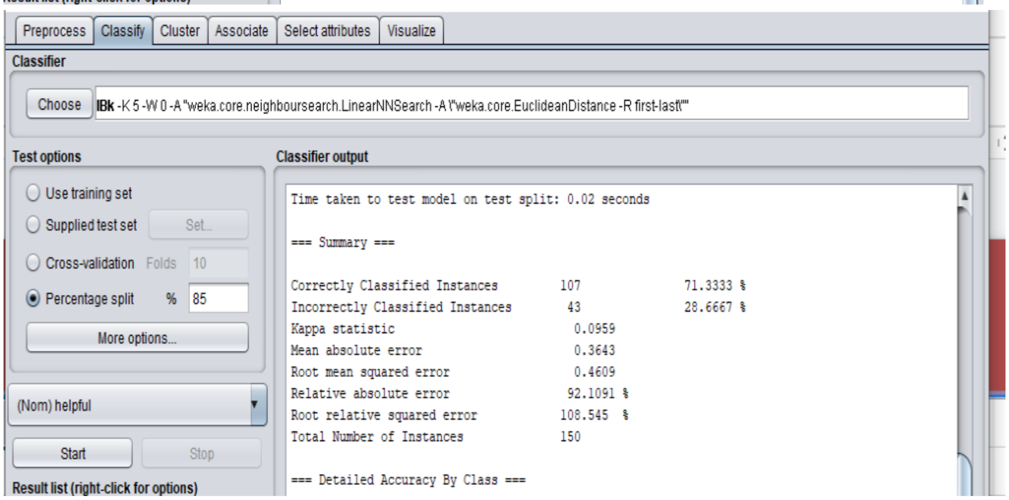
A1 | no

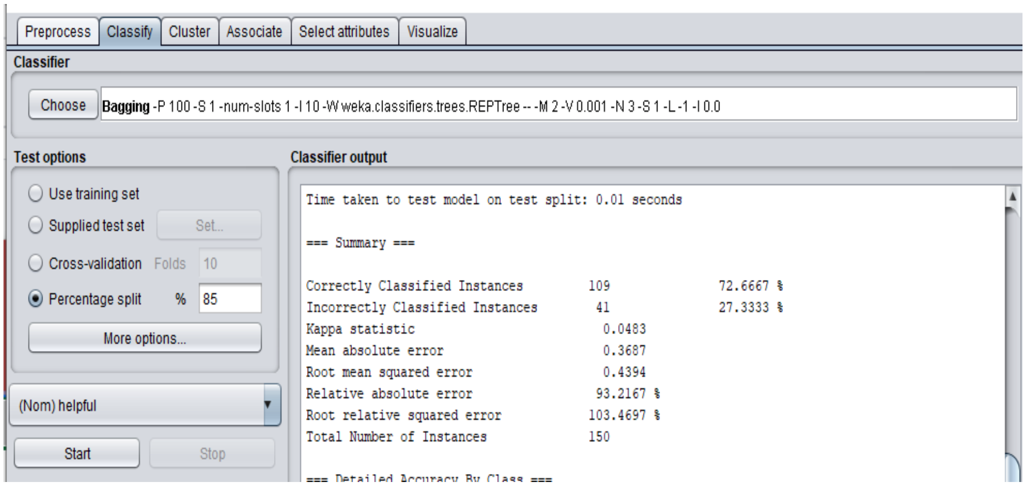
	A	B	C	D	E
1	no	0	0.746667		
2	no	0			
3	yes	1			
4	yes	1			
5	no	0			
6	no	0			
7	yes	1			
8	no	0			
9	no	0			
10	no	0			
11	yes	1			
12	no	0			
13	no	0			
14	yes	1			
15	no	0			
16	no	0			
17	no	0			
18	yes	1			
19	no	0			
20	yes	1			
21	yes	1			
22	no	0			
23	yes	1			
24	no	0			
25	no	0			
26	no	0			

Test 1NN, 3NN and 5NN in Weka the accuracy is 60, 68,71 and 72 separately:









Notes to meet the client:

Red color tries to get more attention from users.

How do the users know about the interface means?

More persuade the users from the design

More instruction to tell the user to interact with the interface of the browser extension
Directly tell the helpfulness of the key word to the users
Confirmed the objectives of the project and achieved most of it