Detecting User Engagement Using Mouse Tracking Data:
Project Specification

David Saunders (910995)

May 2020

Abstract

This project specification reviews relevant materials and the background of my project. The motivation and aims of the project are explained, and a comprehensive plan of work for the summer is present.

Contents

1 Motivation

Crowd-sourcing marketplaces like Amazon's Mechanical Turk are popular services that provide a method for researchers to get participants to complete human intelligence tasks [1]. A common use for this technology is to label data for use in training for machine learning algorithms [2]. However, they can also provide a cheap, scalable method for scientists to gather responses in research. The level of user engagement, attention, and low quality responses can all be issues when gathering data from participants with distributed approaches [3]. The primary motivation of this project is to develop a system of identifying if a crowdsourced user is paying attention during a task.

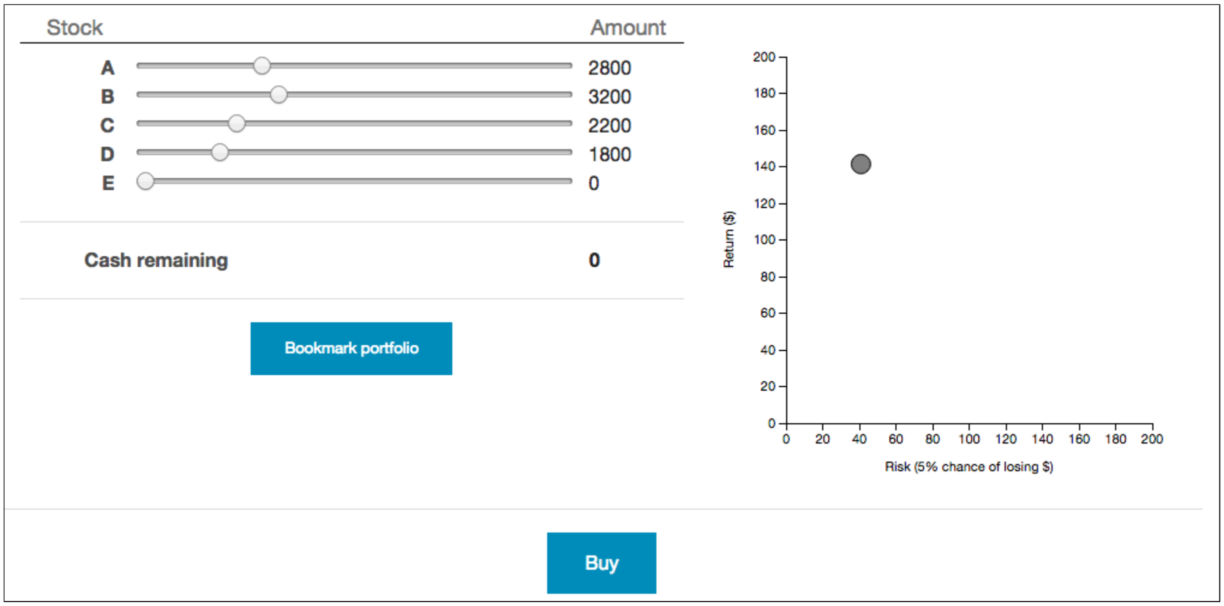
The use of gathering responses using Amazon's Mechanical Turk and other crowdsourcing alternatives are becoming prevalent across disciplines. It is commonly used in conducting clinical research [4] and it is estimated that almost half of all cognitive science research involves the use of crowdsourcing services to collect data samples [5]. The creation of an easy to implement method to measure user engagement would massively help researchers to increase reliability of their research.

Research has been conducted on how the accuracy and attention of crowdsourced tasks can be increased. Methods such as offering financial incentives [6] and engaging a users curiosity [7] have been found to motivate workers into performing better at crowdsourced tasks. Despite the research there is still debate as to which method is superior. If user engagement can be effectively identified then the best method of ensuring user engagement could be found using this method.

Measuring user engagement is well studied in the field of web analytics [8]. However the existing methods of reporting and analysing website data cannot be easily applied to crowdsourced tasks. Characteristics such as session duration and customer satisfaction are used as proxies for engagement. However a longer session doesn't necessarily mean a more engagement user and customer satisfaction is not applicable for crowdsourced tasks. Therefore existing solutions will not work and now methods must be evolved.

2 Introduction

This project uses data from a previous study where participants were asked to perform a simple repetitive tasks [9]. Data was gathered both with a closely monitored lab study, and using a crowd-sourcing website. We assume that participants in the lab study were paying attention, and that crowd sourced participants may or may not be paying attention. The task of this project is to classify crowdsourced participants to determine who were paying attention during the task. We hypothesise that a participants level of attention can be measured from their mouse movement data. The project will propose methods of identifying and quantifying user engagement by using machine learning and visual analytics techniques.



Stock	Amount
A	2800
B	3200
C	2200
D	1800
E	0

Cash remaining **0**

Figure 1: The user interface for the study [9].

Fig 1 is the design of the interface participants of the study interacted with. THe study modelled an investment scenario where participants were asked to maximise profit made over a series of steps. Participants may interact with the stocks sliders to select how much of a stock they would like to buy. The return and risk is shown on the plot on the right of the figure, this plot updates when the sliders are changed in an attempt to inform participants of the stocks volatility. It is hypothesised that participants paying more attention will interact with the sliders more attempting to find the best combination, something their mouse data may reveal.

Table 1: Data collection methods used in the study [9].

Data collection method	Number of participants	Were participants paying attention?
Lab study	18	Yes
Crowdsourced task	370	Unknown

Table 1 shows the two different ways in which data was collected. Only 4.6% of the data was gathered in person, with the vast majority of responses being crowdsourced. The difference in number of participants shows one of the reasons crowdsourcing is popular. It is much easier to crowdsource responses than it is to organise an in person lab study.

2.1 Aims of project

The aims of what I want to achieve in the project will be as follows:

- Visualise, analyse and understand the data.
- Use the data to train machine learning models to classify users by their attention level.
- Determine if there is a link between a users attention and task performance.
- Combine the data and methods from the study data with other crowdsourced datasets to create a more robust model.
- Publicly share all data and findings of this project.

Here I detail how I will achieve each aim, and describe the components of the project that I will need to complete.

To visualise and understand the data I will use tools like Tableau to create basic plots such as histograms. If any interesting correlation or information is found in the data at this stage it can be more fully explored in depth later in the project.

Semi supervised learning techniques can be used to classify which of the crowdsourced participants were paying attention. The idea of semi-supervised learning is combining labelled and unlabelled data to improve the learning behaviour of a machine learning technique [10]. This will be used in this project as only a subset of the data is labelled, where the lab study results are labelled as paying attention and crowdsourced results are unlabelled.

To see if a users attention influences their task performance relies on me being able to extract user attention information from the mouse data. If I am unable to do so I could attempt to use just the lab study data, however there is only a small number of datapoints, and not enough to make a significant

Other datasets can be researched and explored. Similar datasets do exist online and are publicly available [11]. Overfitting is a big issue with any machine learning technique, especially when there is not a large amount of data for training [12]. In an attempt to avoid this I can increase the quantity of data available to me by incorporating other sources and hopefully create a more robust model.

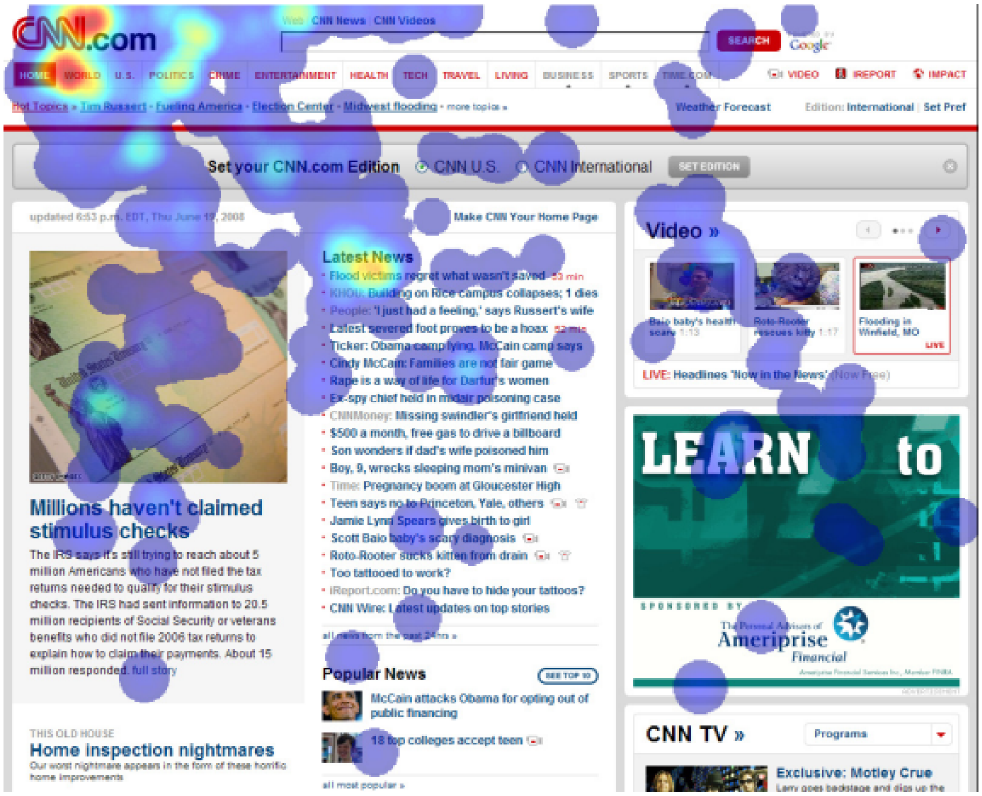
To publicly share the result of this project I will upload any work completed for it in a GitHub repository. Sharing data and results with the Data Science community is an important part of scientific research as it allow others to peer review and reproduce my results [13].

3 Background Research

I have introduced the background as to why it is important to detect user engagement. Now the existing literature will be examined to see how others have proposed to solve this problems and explore any methods that may be used in this project.

3.1 Eye tracking

Non-verbal information such as eye tracking may be used to detect user's level of engagement [14]. Vision is one of the most powerful human senses so it has the potential to give a good measure of user engagement. The methodology of eye tracking is that we move our eyes to focus on particular areas that we want to see in more detail, and divert our attention to that area [15]. Thus tracking a user's gaze can provide insight into which part of a system they're engaged with, and how much so.


CNN.com
Web | CNN News | CNN Videos
SEARCH
Google
HOME
WORLD
U.S.
POLITICS
CRIME
ENTERTAINMENT
HEALTH
TECH
TRAVEL
LIVING
BUSINESS
SPORTS
TIME.COM
VIDEO
iREPORT
IMPACT
Hot Topics » Tim Russert · Fueling America · Election Center · Midwest flooding · more topics »
Weather Forecast
Edition: International | Set Pref
Set your CNN.com Edition
CNN U.S.
CNN International
SET EDITION
Make CNN Your Home Page
Video »
Baio baby's health scare 1:13
Roto-Rooter rescues kitty 1:17
Flooding in Winfield, MO
LIVE
LEARN to
SPONSORED BY
The Personal Advisors of Ameriprise Financial
Ameriprise Financial Services Inc., Member FINRA
CNN TV »
Programs
SEE TOP 10
all news from the past 24hrs »
all most popular »

updated 6:53 p.m. EDT, Thu June 19, 2008

Millions haven't claimed stimulus checks

The IRS says it's still trying to reach about 5 million Americans who have not filed the tax returns needed to qualify for their stimulus checks. The IRS had sent information to 20.5 million recipients of Social Security or veterans benefits who did not file 2006 tax returns to explain how to claim their payments. About 15 million responded. full story

THIS OLD HOUSE

Home inspection nightmares

Our worst nightmare appears in the form of these horrific home improvements

Latest News

- Flood victims regret what wasn't saved 53 min
- KHOU: Building on Rice campus collapses; 1 dies
- People: 'I just had a feeling,' says Russert's wife
- Latest severed foot proves to be a hoax 52 min
- Ticker: Obama camp lying, McCain camp says
- Cindy McCain: Families are not fair game
- Rape is a way of life for Darfur's women
- Ex-spy chief held in midair poisoning case
- CNNMoney: Missing swindler's girlfriend held
- $500 a month, free gas to drive a billboard
- Son wonders if dad's wife poisoned him
- Boy, 9, wrecks sleeping mom's minivan
- Time: Pregnancy boom at Gloucester High
- Teen says no to Princeton, Yale, others
- Jamie Lynn Spears gives birth to girl
- Scott Baio baby's scary diagnosis
- Roto-Rooter sucks kitten from drain
- Too tattooed to work?
- iReport.com: Do you have to hide your tattoos?
- CNN Wire: Latest updates on top stories

Popular News

McCain attacks Obama for opting out of public financing

18 top colleges accept teen

LIVE: Headlines 'Now in the News' (Now Free)

Exclusive: Motley Crue

Larry goes backstage and digs up the

Figure 2: Heatmap showing popular locations of users eyes on a webpage [16].

Eye tracking data can be used to show user interface elements that users focus their attention on as shown in Fig 2. From this researchers were able to predict the amount of attention elements of the page would receive. By observing what parts of an interface users are interacting with we can determine what a user is engaging with [16]. Eye-tracking has been used, and found success novel applications such as recording the engagement of users when playing a game. Tracking users eye movements helped game designers understand how users can recognise interactable game objects and could be used to investigate problematic game design issues [17].

Eivazi and Bednarik extracted features from a users eye tracking data to determine their cognitive state in a problem solving exercise [18]. Features such as "mean fixation duration" and "total path distances" were engineered, and users were split into classes based on their performance. Given a user feature set and their performance class it was able to classify their cognitive state during the task with a 87.5% accuracy with a support vector machine.

Szafir and Mutlu identified a plethora of verbal and non-verbal behavioural cues used by teachers in an educational setting, with gaze being identified as one of them [19]. The behavioural cues could not be recorded directly by a computer, instead EEG signals measured from a headset were used to measure engagement.

Eye tracking is not however a perfect solution and its limitations have been well documented. Track subjects eyes with a good degree of accuracy requires the use of expensive, intrusive equipment that frequently needed recalibrating [20].

3.2 Mouse cursor tracking

Research has also found that there is a correlation between a user's gaze and their cursor position. The position can be considered a "poor man's eye tracker" as it has been found that eye gaze match mouse position 69% of the time [21]. Mouse movement data can be collected without the drawbacks of eye tracking and with more automatic methods, meaning more data can be collected, and on a larger scale [22]. Therefore it can be said that mouse data can be used as a good alternative to eye tracking data.

By using mouse data it is possible to unobtrusively record a user's normal use of a web browser without disturbing their experience [23]. It has also been found that users tend to follow the text they are reading with the mouse cursor [24]. It can be determine what paragraph of a page was being read with an accuracy of 79% by using mouse cursor data [25].

Other methodologies explored ways of classifying user engagement from eye and cursor data, however it is also possible to predict users attention and user frustration in complex webpages [26]. Not all studies agree that mouse cursor is always a good approximation for eye data. Hauger et al found that distinct cursor behaviour exists depending on the task, and that the relationship between eye gaze and mouse position is more nuanced than measuring only mouse data [27].

3.3 Text classification

Text classification is an increasingly popular task in the field of data mining. This typically involves labelling a document into categories by analysing their text content. Examples of this include new filtering where articles are assigned a category or spam filtering where emails can be classified as spam [28].

Nigam et al shows how a the use of Expectation-Maximization and naive Bayes classifier can be used to classify text documents into categories by looking at the distribution of words in a document [29]. A classifier is trained on labelled data which is then used to label unlabelled data. This method of combining supervised and unsupervised learning was found to work better than solely using labelled data. This is especially important with text classification as unlabelled data is much more plentiful than labelled data.

In this paper a document is defined as an "ordered list of word events", however their approach could be applied to other forms of data. We can define a users mouse data as an ordered lists of mouse events. Therefore a similar approach could be used in this project where the frequencies of mouse events are recorded and used in combination with a classification algorithm.

3.3.1 N-Grams

The method by Nigam et al would be considered an unigram model of natural language processing. An n-grams is a series of n words, with unigrams being one word, and bigrams being two word pairs [30].

Different n-grams can be combined together to better understand the complexities of a text document. A mixture of unigrams, bigrams, and trigrams can extract different levels of text complexity and perform well with document classification [28].

4 Description of the project

The high level aims of the project have been given, this section will more concretely detail what work must be done.

4.1 Components of project

The component's can be separated into individual sections, however there is still an order that they must be completed in.

The data must first be extracted from the JSON format that its in to a more tabular format. The project will then consist of researching different algorithms and methods and attempting to implement them on the project data.

Different methods of data extraction or feature engineering may be explored. Any results from this will then be analysed, visualised, and evaluated. These stages can be repeated multiple times as different methods are explored. After multiple different approaches have been explored the results should be properly documented and compiled in my dissertation.

Feature engineering can be approached in a similar was as Eivazi and Bednarik [18]. Features such as total time, time for each step, total mouse distance, mouse movement speed, and time spend on different elements can be extracted. Each of these may be systematically investigated, modified, and evaluated until useful data is engineered. If there is too much variance between users I can attempt to split them into groups based on their task performance. A certain combination of features may revel attention levels in the badly performing group but the approach may not be successful with the well performing group.

The purpose of crowdsourcing responses is to increase the number of participants in the study in an attempt to get a wider spread of data without bias. As a result there are many more crowdsourced responses than lab study responses, as shown in Table 1. Therefore methods of dealing with data with imbalanced classes are another component of work that must be completed. If this is not addressed the the sparseness of data from users definitely engaged in the task may lead to erroneous results when classifying crowdsourced data [31]. The sampling methods of over and under sampling data can solve this problem. Oversampling is where new data are created from the sparse class and with undersampling data is removed from the abundant class [32].

4.2 Description of data

The first component of the project has been completed, and data has been extracted form JSON format to a csv format.

Table 2: The first 5 records of results of the crowdsourced task.

event_type	target	time	x	y	step	turkId
mousedown	alloc-slider-1	0	477	405	1	A35YFAFWP33C70
mouseup	alloc-slider-1	0.111	478	405	1	A35YFAFWP33C70
click	alloc-slider-1	0.111	478	405	1	A35YFAFWP33C70
mousedown	alloc-slider-1	1.516	479	405	1	A35YFAFWP33C70
mousedirchange	alloc-slider-1	2.395	543	403	1	A35YFAFWP33C70
mousedirchange	alloc-slider-1	3.161	594	402	1	A35YFAFWP33C70
mouseup	alloc-slider-1	5.048	514	407	1	A35YFAFWP33C70
click	alloc-slider-1	5.048	514	407	1	A35YFAFWP33C70
mousedown	alloc-slider-2	5.461	494	441	1	A35YFAFWP33C70
mouseup	alloc-slider-2	5.513	494	441	1	A35YFAFWP33C70

Table 2 shows us the features of the data. The lab study data and the crowdsourced data have the same schema, except lab results have a different ID field.

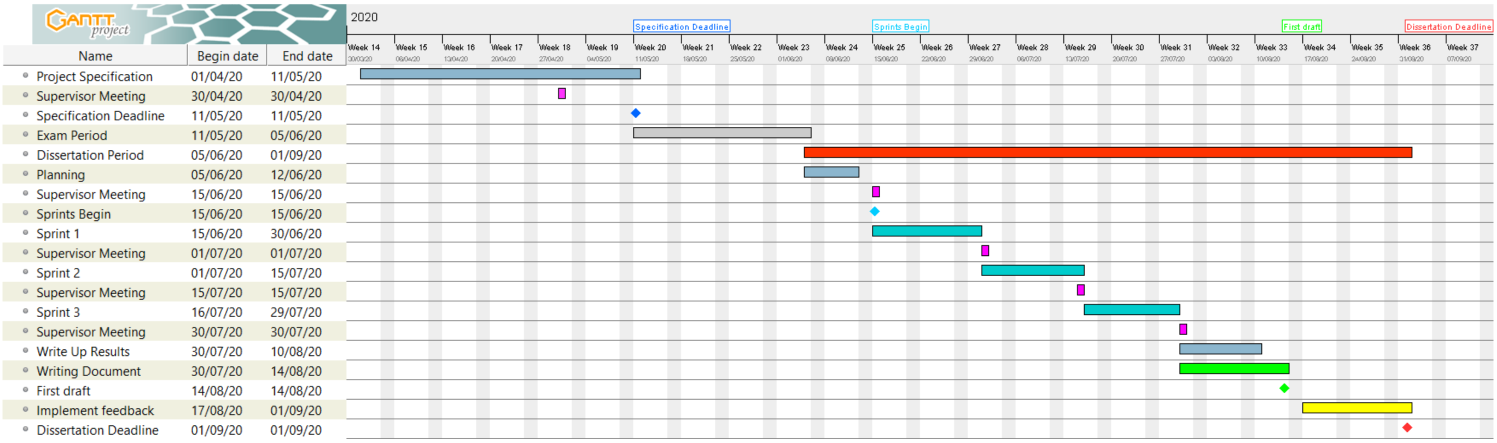
The target field shows us which element in Fig 1 a participant is interacting with and event_type details the type of interaction. Time field shows the time taken in seconds since the first recorded mouse event. We can hypothesise that participants with a shorter time paid less attention than a participant who took much longer, thinking about their actions more. The x and y fields show the location of the mouse and step shows which stage of the task, from 1 to 5, a participant was in.

5 Project plan

The different components of the work have been explained above. This section will specify the timeframe and the order in which the modules will be carried out.

5.1 Development methodology

I will be using an agile methodology as it will allow flexibility of my project and the iterative nature should help me to constantly improve it [33]. Scrum will be used as the short scrum periods will encourage bursts of development over the long summer period [34].



Name	Begin date	End date
Project Specification	01/04/20	11/05/20
Supervisor Meeting	30/04/20	30/04/20
Specification Deadline	11/05/20	11/05/20
Exam Period	11/05/20	05/06/20
Dissertation Period	05/06/20	01/09/20
Planning	05/06/20	12/06/20
Supervisor Meeting	15/06/20	15/06/20
Sprints Begin	15/06/20	15/06/20
Sprint 1	15/06/20	30/06/20
Supervisor Meeting	01/07/20	01/07/20
Sprint 2	01/07/20	15/07/20
Supervisor Meeting	15/07/20	15/07/20
Sprint 3	16/07/20	29/07/20
Supervisor Meeting	30/07/20	30/07/20
Write Up Results	30/07/20	10/08/20
Writing Document	30/07/20	14/08/20
First draft	14/08/20	14/08/20
Implement feedback	17/08/20	01/09/20
Dissertation Deadline	01/09/20	01/09/20

Figure 3: Gantt chart showing the planned timeline and milestones of the project.

Fig 3 shows the sprints I will be undertaking. The Gantt Chart was created with the free software Gantt Project [35].

A sprint will begin with a supervisor meeting where the previous work, and the plan for the next sprint will be discussed. There will be three sprints in total, with each sprint being a fortnight long. Each sprint will consist of researching a method such as a certain machine learning algorithm or different feature engineering techniques I may be able to use in my project.

Time will be spent modifying and implementing the method so that it may be used in the context of this project. The latter half of a sprint will consist of analysing and visualising the results of the method and writing these down roughly in the dissertation document.

The effectiveness of the method will be evaluated and I will attempt to fix any failures or shortcomings of the method. For example I may research variants of artificial neural networks that show promising potential. The sprint plan would allow me to spend time implementing this network with my data and experiment with the parameters.

5.2 Risk analysis

When creating a project there is always potential risks that the project might encounter and hinder its chances of success. In order to prepare and to hopefully avoid these risks I will now list and analyse the risks of my project.

Each risk is explained with the likelihood of the risk occurring and impact to the project the risk would have. A mitigation plan is created in an attempt to prevent the risk from happening, and a contingency plan is made so I can be prepared if the risk does occur. The likelihood of a risk occurring is measured from low to high. Low likelihood risks are unlikely to happen and high is very probable to happen. Impact is measured on the same scale. A low impact risk will have little influence, where as a high project would massively effect it. Below I have listed and analysed the risks and have ordered them from potentially the most dangerous to least dangerous.

1. **Risk:** *Unrealistic time plan and poor time management.*

   Likelihood and Impact: Medium likelihood, Medium Impact

   Explaination If my time is spent poorly then I could not have a piece of work finished for the submission deadline, or the work may not represent the best of my abilities.

   Mitigation: Create work schedule and stick to it. A work schedule and plan for the summer has been created in this document which I aim to follow.

   Contingency: If I am unable to stick to my work schedule, I must adapt my approach to work and create an undated, more realistic schedule.

2. **Risk:** *Coronavirus affects me or a close family member, negatively effecting my work.*

   Likelihood and Impact: Medium likelihood, High Risk

   Explaination Coronavirus is very contagious. In spite of protective measures it is still likely that I may become infected with the virus.

   Mitigation: Stay safe indoors during the quarantine to keep everyone safe and mitigate any risks of me becoming infected.

   Contingency: Inform the University as soon as any negatively situation develops so that alternative assessments can be organised.

3. **Risk:** *No correlation between attention and mouse tracking data can be found.*

Likelihood and Impact: Medium likelihood, High impact

Explaination: The project will involve the use of many methods to find a link between mouse tracking data and user attention. However the number of total data samples are in the hundreds so there is not a massive amount of information to draw conclusions from. It is possible that after all methods have been exhausted no correlation is ever discovered, or simply doesn't exist.

Mitigation: Attempt as many different methods of classification early before writing in depth about them.

Contingency: If no insights can be gained from the given dataset, I will explore other similar datasets and attempt to find correlations there. I will then attempt to apply findings from other datasets to the original dataset.

4. **Risk:** *Coronavirus has a greater impact on Swansea University and effects my available support and deadlines.*

   Likelihood and Impact: Low likelihood, Low impact

   Explaination: The virus has already shut down in person teaching and with the UK in lockdown it is unlikely the situation will become vastly different.

   Mitigation: Keep informed with the University College of Science and supervisor to any news effecting the University.

   Contingency: Keep updated with the situation and follow whatever advice is recommended from the university.

6 Conclusion

In this report the motivation of why there is a need to detect user engagement in increasingly popular crowdsourcing services. The problem of detecting user engagement in crowdsourced responses has been explained and the need for a new solution is outlined.

The existing methods of eye, and mouse tracking have been researched. It was found that mouse data has many of the benefits of eye tracking data with none of the drawbacks, making it ideal for use in user engagement.

A detailed plan of the project was shown, with a detailed breakdown of the timeline, and components of the project shown with a Gantt chart. Risk analysis of the project was completed so that no unexpected incident will disrupt the project.

References

[1] Gabriele Paolacci, Jesse Chandler and Panagiotis G Ipeirotis. “Running experiments on amazon mechanical turk”. In: *Judgment and Decision making* 5.5 (2010), pp. 411–419.

[2] Joseph Chee Chang, Saleema Amershi and Ece Kamar. "Revolt: Collaborative crowdsourcing for labeling machine learning datasets". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 2334–2346.

[3] Panagiotis G Ipeirotis, Foster Provost and Jing Wang. "Quality management on amazon mechanical turk". In: *Proceedings of the ACM SIGKDD workshop on human computation*. 2010, pp. 64–67.

[4] Jesse Chandler and Danielle Shapiro. "Conducting clinical research using crowdsourced convenience samples". In: *Annual review of clinical psychology* 12 (2016).

[5] Neil Stewart, Jesse Chandler and Gabriele Paolacci. "Crowdsourcing samples in cognitive science". In: *Trends in cognitive sciences* 21.10 (2017), pp. 736–748.

[6] Chien-Ju Ho et al. "Incentivizing high quality crowdwork". In: *Proceedings of the 24th International Conference on World Wide Web*. 2015, pp. 419–429.

[7] Edith Law et al. "Curiosity killed the cat, but makes crowdwork better". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 4098–4110.

[8] Eric T Peterson and Joseph Carrabis. "Measuring the immeasurable: Visitor engagement". In: *Web Analytics Demystified* 14 (2008), p. 16.

[9] Thomas Torsney-Weir et al. "Risk Fixers and Sweet Spotters: a Study of the Different Approaches to Using Visual Sensitivity Analysis in an Investment Scenario". In: *EuroVis 2018 - Short Papers*. Ed. by Jimmy Johansson, Filip Sadlo and Tobias Schreck. The Eurographics Association, 2018. ISBN: 978-3-03868-060-4. DOI: 10.2312/eurovisshort.20181089.

[10] Xiaojin Zhu and Andrew B Goldberg. "Introduction to semi-supervised learning". In: *Synthesis lectures on artificial intelligence and machine learning* 3.1 (2009), pp. 1–130.

[11] Human Computation. *Workers Browser Activity in CrowdFlower Tasks*. Oct. 2017. URL: https://www.kaggle.com/humancomp/worker-activity-crowdflower (visited on 03/05/2020).

[12] Tom Dietterich. "Overfitting and undercomputing in machine learning". In: *ACM computing surveys (CSUR)* 27.3 (1995), pp. 326–327.

[13] Jeremy P. Birnholtz and Matthew J. Bietz. "Data at Work: Supporting Sharing in Science and Engineering". In: *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work*. GROUP '03. Sanibel Island, Florida, USA: Association for Computing Machinery, 2003, pp. 339–348. ISBN: 1581136935. DOI: 10.1145/958160.958215. URL: https://doi.org/10.1145/958160.958215.

[14] Divesh Lala et al. "Detection of social signals for recognizing engagement in human-robot interaction". In: *arXiv preprint arXiv:1709.10257* (2017).

[15] Andrew T Duchowski. “Eye tracking methodology”. In: *Theory and practice* 328.614 (2007), pp. 2–3.

[16] Georg Buscher, Edward Cutrell and Meredith Ringel Morris. “What do you see when you’re surfing? Using eye tracking to predict salient regions of web pages”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2009, pp. 21–30.

[17] Tony Renshaw, Richard Stevens and Paul D Denton. “Towards understanding engagement in games: an eye-tracking study”. In: *On the Horizon* (2009).

[18] Shahram Eivazi and Roman Bednarik. “Predicting problem-solving behavior and performance levels from visual attention data”. In: *Proc. Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI*. 2011, pp. 9–16.

[19] Daniel Szafir and Bilge Mutlu. “Pay attention! Designing adaptive agents that monitor and improve user engagement”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2012, pp. 11–20.

[20] Daniel C Richardson and Michael J Spivey. “Eye tracking: Characteristics and methods”. In: *Encyclopedia of biomaterials and biomedical engineering* 3 (2004), pp. 1028–1042.

[21] Lynne Cooke. “Is the Mouse a” Poor Man’s Eye Tracker”?” In: *Annual Conference-Society for Technical Communication*. Vol. 53. 2006, p. 252.

[22] Urška Demšar and Arzu Çöltekin. “Quantifying gaze and mouse interactions on spatial visual interfaces with a new movement analytics methodology”. In: *PloS one* 12.8 (2017).

[23] Jeremy Goecks and Jude Shavlik. “Learning users’ interests by unobtrusively observing their normal behavior”. In: *Proceedings of the 5th international conference on Intelligent user interfaces*. 2000, pp. 129–132.

[24] Chen-Chung Liu and Chen-Wei Chung. “Detecting mouse movement with repeated visit patterns for retrieving noticed knowledge components on web pages”. In: *IEICE transactions on information and systems* 90.10 (2007), pp. 1687–1696.

[25] David Hauger, Alexandros Paramythis and Stephan Weibelzahl. “Using browser interaction data to determine page reading behavior”. In: *International Conference on User Modeling, Adaptation, and Personalization*. Springer. 2011, pp. 147–158.

[26] Vidhya Navalpakkam and Elizabeth Churchill. “Mouse tracking: measuring and predicting users’ experience of web-based content”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, pp. 2963–2972.

[27] Jeff Huang, Ryen White and Georg Buscher. “User see, user point: gaze and cursor alignment in web search”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012, pp. 1341–1350.

[28] Charu C Aggarwal and ChengXiang Zhai. “A survey of text classification algorithms”. In: *Mining text data*. Springer, 2012, pp. 163–222.

[29] Kamal Nigam et al. “Text classification from labeled and unlabeled documents using EM”. In: *Machine learning* 39.2-3 (2000), pp. 103–134.

[30] James Martin Daniel Jurafsky. *Speech and Language Processing*. 2009.

[31] Miroslav Kubat, Stan Matwin et al. “Addressing the curse of imbalanced training sets: one-sided selection”. In: *Icml*. Vol. 97. Citeseer. 1997, pp. 179–186.

[32] Chris Drummond, Robert C Holte et al. “C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling”. In: *Workshop on learning from imbalanced datasets II*. Vol. 11. Citeseer. 2003, pp. 1–8.

[33] Kent Beck et al. “Manifesto for agile software development”. In: (2001).

[34] Ken Schwaber. “Scrum development process”. In: *Business object design and implementation*. Springer, 1997, pp. 117–134.

[35] GanttProject Team. *GanttProject - Free project scheduling and management app for Windows, OSX and Linux*. Apr. 2020. URL: `https://www.ganttproject.biz/` (visited on 03/05/2020).