**SECTION A: Project Definition**

**FOR UNDERGRADUATE & TAUGHT POSTGRADUATE ONLY**

**Complete the following table with full and relevant information relating to your research.**

|  |  |
| --- | --- |
| Student Name | Mark Baber |
| Student Number | 17076749 |
| Student E-mail Address (please use University e-mail) | [17076749@students.southwales.ac.uk](mailto:17076749@students.southwales.ac.uk) |
| Name of Principal Project Supervisor | Dr Paul Jarvis |
| Project Title | To determine whether there is a link between the sentiment of a TV show and the viewer ratings. |
| Briefly describe the project, being sure to identify any aspects that are relevant to the Ethical Evaluation in Section B.  NOTE: A project determined to be High Risk will need to include additional information in Section B to fully-specify the risks and mitigations. | In this project I am going to use sentiment analysis to determine if there is a link between the sentiment of a television episode and its review score.  This will entail collecting data from public domain websites which allow data mining and websites which are specifically for data science with open source data available for all users.  I will also carry out some investigation into collecting data from public domain websites i.e. Twitter, IMDB and Rotten Tomatoes depending on the scope of this project.  Twitter - For using Twitter, I would need to sign up for a development account (free) which as stated by Twitter, “enable programmatic analysis of Tweets back to the first Tweet in 2006”. Using their free API will not allow you to take any personal information from Twitters users.  IMDB – If using IMDB they allow personal and non-commercial use of their datasets which can also be downloaded locally, which stops users from scrapping data from their live site.  Rotten Tomatoes – If using Rotten Tomatoes, I need to sign up for the development program which allows me to use their API, this API will give me access to two types of data:   * Critic and Audience Scores * Critic Reviews   With these types of data it would help me a lot through my research.  If I do have time to collect data, I will make sure all users have an anonymous but unique id, deducting any names. The data will then be collected together and be broken down word for word, which will make identifying a user very difficult.  This research may include data from the following sources:   * Children under 16, as all the terms of service and privacy policy of all websites mentioned allow users as young as 14 to sign up and register. None of these will be targeted individually but will be used more of a mass of words in no order. Which will make it harder to identify an individual among the mass of words/data collected. * Adults with learning difficulties, whilst not specifically targeted in this research, there isn’t anything in the terms of service or the privacy policy about what data companies collect on users who have a disability. Again, as stated above, if there are any users which fall under this category they will not be targeted individually.   This research will not collect any data from the following sources:   * Adults or children with diagnosed mental illness/terminal illness/dementia/in a residential care home. * Adults or children in emergency situations. * Adults or children with limited capacity to consent * those who are “dependent” on others   This research does not require any of the following:   * Requires NHS ethical approval * Requires a Human Tissue Act license. * Involves “covert” procedures as in covert observation studies.   This research does not include the following areas:   * Anything considered “sensitive”. * Significant psychological stress or anxiety or produce humiliation or cause more than fleeting harm / negative consequences beyond the risks encountered in the normal life of the participants (and where the potential for fleeting “harm” is clearly detailed in the participant information sheet). If in doubt regarding definition of the above terminology, please contact the research governance office. * Administration of drugs, placebos or other substances (such as food substances or vitamins) as part of this study. * Invasive procedures (not limited to blood sampling, collection of biological samples, or passing current through a participant’s body, etc.). * Inducements to participate in the study. * Recruitment of serving prisoners or serving young offenders via Her Majesty’s Prison & Probation Service. |
| Please add an explanation of your study in plain English, with focus on any parts of your study which involve human participants. No more than 100 words. This is to help the Faculty Research Ethics Committee (FREC) to understand the project. | To collect and analyse the scripts from a TV show and use the programming language R to analyse the scripts. This will be done with R packages which allow the user to organise, tidy and analyse the data.  When presenting my finding at the end I will do so ethically and respectfully and doesn’t involve any of the issues highlighted below.  The scripts can be found online, and the views and ratings can be found online (and are in the public domain), and to see if there is a link between the main sentiment of an episode (sad, confusing) and the number of views.  When collecting data, I will make sure all users have an anonymous but unique id, deducting any names. i.e. - u\_17546  When processing the data all words will be tokenized using the bag of words method, meaning words will be identifiable under an ID (as stated above) |

**SECTION B: Ethical Evaluation**

**FOR UNDERGRADUATE & TAUGHT POSTGRADUATE ONLY**

Consider the following points to determine the level of ethical risk your research presents:

* Involves those who are considered vulnerable such as:
* Children under 16.
* Adults with learning difficulties.

Unless in an accredited setting, accompanied by a carer or professional with a duty of care.

* Involves those who are considered highly vulnerable such as:
* Adults or children with diagnosed mental illness/terminal illness/dementia/in a residential care home.
* Adults or children in emergency situations.
* Adults or children with limited capacity to consent
* Involves those who are “dependent” on others (such as teacher or lecturer to student). Unless in an accredited setting associated with normal working conditions or routines and within normal operating hours, such as a cultural institution, pre-school, school, or youth club where the research is carried out as part of professional practice such as curriculum development.
* Requires full NHS ethical approval via the Integrated Research Application System.
* Requires a Human Tissue Act license.
* Involves “covert” procedures as in covert observation studies.
* Involves anything considered “sensitive”. For example, does not carry a risk of those involved disclosing information which compromises the research (e.g., illegal activities; activities where moral opinion may differ, potential professional misconduct – work errors).
* Induces significant psychological stress or anxiety or produce humiliation or cause more than fleeting harm / negative consequences beyond the risks encountered in the normal life of the participants (and where the potential for fleeting “harm” is clearly detailed in the participant information sheet). If in doubt regarding definition of the above terminology, please contact the research governance office.
* Involves administration of drugs, placebos or other substances (such as food substances or vitamins) as part of this study.
* Involves invasive procedures (not limited to blood sampling, collection of biological samples, or passing current through a participant’s body, etc.).
* Offers any financial inducements to participate in the study.
* Intends to recruit serving prisoners or serving young offenders via Her Majesty’s Prison & Probation Service.

For your course, there may be specific requirements in **addition** to these, depending on the nature of the subject and how your project is assessed. You must also complete those requirements.

If **none** of the 12 points above apply, then the research can be considered **Low Risk**, *unless your course identifies additional criteria relevant to your subject that would render it High Risk*. This Section is then signed off by yourself and your supervisor and held on file for review by FREC.

If **any** of the 12 points applies, then the research is considered **High Risk** and students must bring the matter to the attention of their research supervisor immediately. **Research cannot then commence until mitigations for the risk are agreed by FREC**. Seek advice from your Supervisor, who can help you identify mitigations of the risk or redesign as a Low Risk project.

**All students must complete the section below, in collaboration with their supervisor.**

Please strike through the statement that **does not** apply.

1. An ethics review has been completed, and the project has been identified as Low Risk.
2. ~~An ethics review has been completed, and a High Risk was identified. I agree to explain how they may be mitigated below and agree to abide by any conditions identified at this stage, by my Project Supervisor, the School or the Faculty. I understand that High Risk projects can only proceed with approval from the Faculty Research Ethics Committee.~~

|  |
| --- |
| Issues: (Include as much information as possible to help FREC members to understand the issues. Extend onto additional pages as necessary.)  None of the above apply to my study.  The data gathered will be public domain, and not linked to any person or organisation that can easily identified. This will be done by having all users’ comments linked to a unique id and will be combined with a mass amount of data (Not following one single person’s comments) so would make it harder to identify a single user. |
| Proposed mitigations: (Include as much information as possible to help FREC members to understand the mitigations. Extend onto additional pages as necessary.)  When collecting data from websites, all names would be anonymised with a unique id, i.e. - u\_17546 so we’re able to see other reviews and find patterns. |
| Student’s Signature:  Date: |
| **Supervisor’s statement:** I have ensured due diligence and accountable decision making by the student. I have sought appropriate advice where required to support my judgment in this.  Supervisor’s Signature:  Date: |
| **Any false or mis-represented information contributing to this Ethical Evaluation, including attempting to pass off a High-Risk project as a Low Risk project, is subject to the Student Misconduct Regulations and may also have legal repercussions.** |

Both signatures are **required** for all projects, both Low Risk and High Risk.