# 

# **UREC 1 RESEARCH ETHICS REVIEW FOR STUDENT RESEARCH WITH NO HUMAN PARTICIPANTS OR DIRECT COLLECTION OF HUMAN TISSUES, OR BODILY FLUIDS.**

All University research is required to undergo ethical scrutiny to comply with UK law. The University Research Ethics Policy ([www.shu.ac.uk/research/excellence/ethics-and-integrity/policies](https://www.shu.ac.uk/research/excellence/ethics-and-integrity/policies)) should be consulted before completing this form. The initial questions are there to check that completion of the UREC1 is appropriate for this study. The supervisor will approve the study, but it may also be reviewed by the College Teaching Program Research Ethics Committee (CTPREC) as part of the quality assurance process (additional guidance can be obtained from your College Research Ethics Chair[[1]](#footnote-0)).

The final responsibility for ensuring that ethical research practices are followed rests with the supervisor for student research.

Note that students and staff are responsible for making suitable arrangements to ensure compliance with the General Data Protection Regulations (GDPR), for keeping data secure and if relevant, for keeping the identity of participants anonymous. They are also responsible for following SHU guidelines about data encryption and research data management. Guidance can be found on the SHU Ethics Website [www.shu.ac.uk/research/excellence/ethics-and-integrity](https://www.shu.ac.uk/research/excellence/ethics-and-integrity)

Please note that it is mandatory for all students to only store data on their allotted networked F drive space and not on individual hard drives or memory sticks etc.

This form also enables the University and College to keep a record confirming that research conducted has been subjected to ethical scrutiny. Students should retain a copy for inclusion in their research projects, and a copy should be uploaded to the relevant module Blackboard site.

The form must be completed by the student and approved by supervisor and/or module leader (as applicable). In all cases, it should be counter-signed by the supervisor and/or module leader and kept as a record showing that ethical scrutiny has occurred. Students should retain a copy for inclusion in the appendices of their research projects, and a copy should be uploaded to the module Blackboard site for checking.

Please note that it may be necessary to conduct a health and safety risk assessment for the proposed research. Further information can be obtained from the University’s Health and Safety Website https://sheffieldhallam.sharepoint.com/sites/3069/SitePages/Risk-Assessment.aspx

**ARE YOU COMPLETING THE CORRECT FORM?**

Does this study include collecting data or samples from human participants. YES/NO

Is the secondary data used in this study of a sensitive or contentious nature, or does it allow the identification of individuals or organisations (e.g., companies, school, councils, communities). YES/NO

If you have answered ***YES*** to either of these two questions you must complete a UREC2, 3 or 4 as appropriate.

## **1. General Details**

| **Details** |  |
| --- | --- |
| Name of student | Naveen Ramaiya Selvaraj |
| SHU email address | C3075306@hallam.shu.ac.uk |
| Department/College | Msc Big data Analytics / Sheffield Hallam University |
| Name of supervisor | Sandhya Makkar |
| Supervisor’s email address | S.Makkar@shu.ac.uk |
| Title of proposed research | Predictive Analytics in Social Media - Forecasting Consumer Behavior and Market Trends |
| Proposed start date |  |
| Proposed end date |  |
| Brief outline of research to include, rationale (reasons) for undertaking the research & aims, and methods (max 500 words). | **Overview and Justification**  Social media's explosive expansion has produced a previously unheard-of amount of user-generated data that can provide profound insights into consumer behaviour and new industry trends. It is still difficult to turn this massive amount of unstructured data into forecasts that can be put into practice. It is frequently impossible for traditional analytical methods to identify intricate, non-linear patterns in social media data. In order to increase the precision of predictive analytics in social media environments, this study suggests utilising cutting-edge machine learning models, including Random Forest, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT). These models were chosen because of their special abilities to handle text, sequential, and structured data all of which are frequently present in social media content. The requirement for reliable, scalable prediction models that can provide accurate and timely projections based on social media activity is a significant gap that this research fills. Companies are depending more and more on these data to inform product development, customise marketing, and adapt to changing customer needs. The study's findings could improve business decision-making and help firms remain competitive by adjusting to current consumer trends.  **Research Objective**  Finding out how well different machine learning models such as Random Forest Classifier, SVM, LSTM, and BERT predict consumer behaviour and market trends using social media data is the main goal of this study. The goal of the study is to determine the most effective model for precise, comprehensible forecasts that can aid in strategic decision-making in product development and digital marketing.  **Methods**  **Gathering and Preparing Data**  In order to gather a variety of consumer data, the study will make use of social media tweet datasets that will be extracted from X platform using web scraping techniques from 2017 to 2023, and it will be scraped directly from X using Twitter API. The dataset will include tweet data of four fashion brands namely [H&M](https://x.com/hm), [Gap](https://x.com/Gap), [Uniqlo](https://x.com/UNIQLO_UK) and [Zara](https://x.com/ZARA). This dataset can help analyze the relationship between Twitter advertising and consumer behaviour in the fashion industry. Some of the variables that will be used to extract tweets include - "Used\_Hashtag(s)", "Tweet\_Content", "Post\_Length", "Likes\_Count", "Tweet\_DateTime", "Replies\_Count", "Retweets\_Count", "Image(s)", "video(s)", "GIF(s)", "HaveText", "Tweet\_URL", "Reply\_Content" and "Reply\_URL". To guarantee quality and consistency, data will go through preprocessing procedures like cleaning, tokenisation, and normalisation. While structured data characteristics will be produced for Random Forest and SVM analysis, the textual material will be prepared for analysis using NLP approaches, making it appropriate for the BERT and LSTM models.  **Implementation of the Model**   * Random Forest Classifier is a model that works well with tabular, structured data that exhibits non-linear patterns. It sorts data into groups using a group of decision trees, which makes it reliable and accurate with a low chance of overfitting. * Support Vector Machine (SVM) is a potent classifier that operates by determining the best line dividing classes in the data. It works well in situations when there are distinct class differences, which makes it helpful for classifying sentiment or behaviour. * LSTM is a recurrent neural network model that is perfect for analysing time-series data, including post frequency, trends over time, and behavioural changes. It is made to capture sequential dependencies. * BERT is a transformer-based model that is very good at interpreting textual context, which makes it ideal for sentiment analysis and drawing conclusions from user reviews and comments. It is able to decipher complicated linguistic patterns and nuanced emotions that are frequently found in social media posts.   **Evaluation**  Recall, accuracy, precision, and F1-score will all be used to assess each model's performance. Additionally, since interpretability and computing efficiency are critical for real-time consumer behaviour analysis, they will be taken into account. |

## **2. Research in External Organisations**

| **Question** | **Yes/No** |
| --- | --- |
| 1. Will the research involve working with/within an external organisation (e.g., school, business, charity, museum, government department, international agency, etc.)? | No |
| 1. If you answered YES to question 1, do you have granted access to conduct the research?   *If YES, students please show evidence to your supervisor. PI should retain safely.* |  |
| 1. If you answered NO to question 2, is it because:    1. you have not yet asked    2. you have asked and not yet received an answer    3. you have asked and been refused access.   *Note: You will only be able to start the research when you have been granted access.* |  |

## **Research with Products and Artefacts**

| **Question** | **Yes/No** |
| --- | --- |
| 1. Will the research involve the use of specialist copyrighted documents, films, broadcasts, photographs, artworks, designs, products, programs, databases, networks, processes, existing datasets, or secure data? | Yes |
| 2. If you answered YES to question 1, are the materials you intend to use in the public domain?  *Notes: ‘In the public domain’ does not mean the same thing as ‘publicly accessible’.*   * *Information which is 'in the public domain' is no longer protected by copyright (i.e., copyright has either expired or been waived) and can be used without permission.* * *Information which is 'publicly accessible' (e.g., TV broadcasts, websites, artworks, newspapers) is available for anyone to consult/view. It is still protected by copyright even if there is no copyright notice. In UK law, copyright protection is automatic and does not require a copyright statement, although it is always good practice to provide one. It is necessary to check the terms and conditions of use to find out exactly how the material may be reused etc.*   *If you answered YES to question 1, be aware that you may need to consider other ethics codes. For example, when conducting Internet research, consult the code of the Association of Internet Researchers; for educational research, consult the Code of Ethics of the British Educational Research Association.* | Yes |
| 3. If you answered NO to question 2, do you have explicit permission to use these materials as data?  *If YES, please show evidence to your supervisor.* |  |
| 4. If you answered NO to question 3, is it because:  A. you have not yet asked permission  B. you have asked and not yet received and answer  C. you have asked and been refused access.  *Note: You will only be able to start the research when you have been granted permission to use the specified material.* | **A/B/C** |

1. **Does this research project require a health and safety risk assessment for the procedures to be used?** (Discuss this with your supervisor)

☐ Yes

☒ No

If **YES** the completed Health and Safety Risk Assessment form should be attached. A standard risk assessment form can be generated through the Awaken system (<https://shu.awaken-be.com>). Alternatively if you require more specific risk assessment, e.g. a COSHH, attach that instead.

**Insurance Check**

The University’s standard insurance cover will not automatically cover research involving any of the following:

i) Participants under 5 years old

ii) Pregnant women

iii) 5000 or more participants

iv) Research being conducted in an overseas country

v) Research involving aircraft and offshore oil rigs

vi) Nuclear research

vii) Any trials/medical research into Covid 19

If your proposals do involve any of the above, please contact the Insurance Manager directly ([fin-insurancequeries-mb@exchange.shu.ac.uk](mailto:fin-insurancequeries-mb@exchange.shu.ac.uk)) to discuss this element of your project.

## **Adherence to SHU Policy and Procedures**

| **Ethics sign-off** | |
| --- | --- |
| **Personal statement** | |
| I can confirm that:   * I have read the Sheffield Hallam University Research Ethics Policy and Procedures * I agree to abide by its principles. | |
| **Student** | |
| Name: | Date: |
| Signature: | |
| **Supervisor ethical sign-off** | |
| I can confirm that completion of this form has confirmed that this research does not involve human participants. The research will not commence until any approvals required under Sections 2 & 3 have been received and any health and safety measures are in place. | |
| Name: | Date: |
| Signature: | |
| **Independent Reviewer ethical sign off** (if required to permit publication of findings with supervisor co-authorship). | |
| Name: | Date: |
| Signature: | |

**Please ensure that you have attached all relevant documents. Your supervisor must approve them before you start data collection:**

| **Relevant Documents** | **Yes** | **No** | **N/A** |
| --- | --- | --- | --- |
| Research proposal if prepared previously | **☐** | **☐** | **☐** |
| Any associated materials (e.g., posters, letters, etc.) | **☐** | **☐** | **☐** |
| Health and Safety Risk Assessment Form | **☐** | **☐** | **☐** |

1. College of Social Sciences and Arts – Dr. Antonia Ypsilanti ([a.ypsilanti@shu.ac.uk](mailto:a.ypsilanti@shu.ac.uk) )

   College of Business, Technology and Engineering – Dr. Tony Lynn ([t.lynn@shu.ac.uk](mailto:t.lynn@shu.ac.uk) )

   College of Health, Wellbeing and Life Sciences – Dr. Nikki Jordan-Mahy ([n.jordan-mahy@shu.ac.uk](mailto:n.jordan-mahy@shu.ac.uk) ) [↑](#footnote-ref-0)