

**DUNDALK INSTITUTE OF TECHNOLOGY**  
**School of Informatics & Creative Arts**  
**Ethical Approval Form for Research Projects**

Researcher Name Murthy S Routhula Year 2021-2022 Course M.Sc. Data Analytics

Title of project Studying the Effect of Vectorization Techniques in Mix-Code (Hinglish Language) on Open-Source Data Using Machine Learning and Transfer Learning Methodology.

Name of supervisor/s Dr. Abhishek Kaushik Date 25<sup>th</sup> March 2022  
(if applicable)

*This application is to be completed by the researcher and where appropriate, in conjunction with the project supervisor. The lead researcher/supervisor is responsible for submitting the completed form to the appropriate Research Ethics Committee (details below)*

**Please note: If your submission is incomplete or unclear, your application will be returned to you and your project may be delayed.**

## Section 1

### Type of Researcher

Please tick the appropriate box below to indicate the type of researcher you are:

**Undergraduate**

(Proceed to section 2)

☐

*Completed Ethical Approval forms for undergraduate research should be submitted to the relevant Departmental Research Ethics Committee (DREC).*

[Drec.dcam@dkit.ie](mailto:Drec.dcam@dkit.ie) – Department of Creative Arts Media and Music

[Drec.dcs@dkit.ie](mailto:Drec.dcs@dkit.ie) – Department of Computing Science and Mathematics

[Drec.dvhcc@dkit.ie](mailto:Drec.dvhcc@dkit.ie) – Department of Visual and Human Centred Computing

**Postgraduate**

(Proceed to section 3)

☒

*Completed Ethical Approval forms for Postgraduate research should be submitted directly to the School Research Ethics Committee (SREC).*

[Srec.ica@dkit.ie](mailto:Srec.ica@dkit.ie)

**Staff**

(Proceed to section 3)

☐

*Completed Ethical Approval forms for Staff research should be submitted directly to the School Research Ethics Committee (SREC).*

[Srec.ica@dkit.ie](mailto:Srec.ica@dkit.ie)

## Section 2

Please complete questions 1-4 listed below.

	Human and / or Animal Research	YES	NO
1	<p><b>Does your research involve human participants other than the following<sup>1</sup>?</b></p> <ul style="list-style-type: none"> <li>• Research using exclusively secondary sources.</li> <li>• Research using materials legally accessible to the public that have legal protection, e.g., record of court judgements, data archives.</li> <li>• Research using materials that are publicly accessible and where there is no reasonable expectation for privacy, e.g., books, published third party interviews.</li> <li>• Observations of human behavior in public where (i) those being observed have no reasonable expectation of privacy, (ii) there is no intervention on the part of the researcher nor any interaction between the researcher and those observed, and (iii) individuals are not identifiable in the results.</li> </ul> <p>If 'YES', please complete <b>B and C</b> below.</p>		
2	<p><b>Does your project involve working with animals?</b></p> <p>– If 'YES', please complete <b>B and C</b> below. – Please note that for ethical consideration: 'Animals' are classed as vertebrate animals including cyclostomes and cephalopods (DIRECTIVE 2010/63/EU)</p>		
3	<p><b>Does your project involve working with participants from any of the following categories?</b></p> <ul style="list-style-type: none"> <li>• Minors (under 18 years of age)</li> <li>• People with learning or communication difficulties</li> <li>• Patients</li> <li>• People in custody</li> <li>• People engaged in illegal activities</li> </ul> <p>If 'YES', please complete <b>D</b> below.</p>		
4	<p><b>Does your project have any possible ethical implications other than those outlined in questions 1, 2 and 3?</b></p> <p>If 'YES', please complete <b>E</b> below.</p>		

**A. I consider that this project has no significant ethical implications<sup>2</sup> to be brought through the ICA School Ethics Review Process**  
**Please complete Section 4**

☐

<sup>1</sup> If there is any doubt, researchers should contact the Chair of the SREC.

<sup>2</sup> In determining significant implications, please consider all potential risks attached to this project. If there is any doubt, researchers should contact the Chair of the SREC.

**B.**

- I. Is this study part of a larger project that already has ethical clearance?**

YES / NO

If **YES** please answer question B II.

If **NO** please answer question C.

- II. If this study is part of a larger project that already has ethical clearance, are you proposing any changes to the operational plan already ethically approved?**

YES / NO

If **YES**, please complete *Sections 3 and 4*.

If **NO**, then please provide the project details below and complete *Section 4*.

Title of project with ethical clearance: \_\_\_\_\_

- C. Could this project have ethical implications that should be brought before the appropriate ICA Departmental Ethics Review Committee as it will be carried out with human participants?**

YES / NO

If **YES**, please complete *Sections 3 and 4*.

If **NO**, please complete *Section 4*.

- D. I consider that this project may have ethical implications that should be brought before the School Research Ethics Committee as it will be carried out with human participants in a “vulnerable” category.**

***Please complete Sections 3 and 4***

☐

**All research carried out with human participants in a vulnerable category must be referred by the Departmental Research Ethics Committee to the School Research Ethics Committee for approval.**

- E. Could this project have ethical implications, other than those previously outlined, that should be brought before the appropriate ICA Departmental Ethics Review Committee?**

YES / NO

If **YES**, please complete *Sections 3 and 4*.

If **NO**, please complete *Section 4*.

## Section 3

### 3.1 Application Form Checklist

Please complete Section 3 and provide additional information as attachments.

My application includes the following documentation:	INCLUDED (mark as YES)	NOT APPLICABLE (mark as N/A)
Recruitment advertisement		N/A
Participant Information Leaflet		N/A
Participant Informed Consent form		N/A
Questionnaire/Survey		N/A
Interview/Focus Group Questions		N/A
Debriefing material		N/A
Evidence of approval to gain access to off-site location		N/A
Ethical approval from external organizations. If ethical approval from external organizations is pending give details below		N/A
Details: N/A		

### 3.2 Project Details

a) Lay description (Maximum 200 words)

Please outline, in terms that any non-expert would understand, what your research project is about, including what participants will be required to do. Please explain any technical terms or discipline-specific phrases.

This project focus on the different stat-of-the-art (SOTA) Vectorization techniques and study its effect on real life natural language problem such as automating the Sentiment Analysis in the Mix code language. Mix code language is the combination of two language where the essence of one language is used for writing and the other language is used for sense making. For an example, “*muje ye accha laga tha*”. The script is written in English but the sense making in in Hindi which is “*I liked it*”. These scenarios are very common in the countries like India, Nepal etc.

In this project, I am working on YouTube Channels comments based on Hinglish (Hindi + English) language and their sentiments. I will carry out the Sentimental Analysis on those comments using Natural Language Processing and Parametric and Non-Parametric Approaches of Machine Learning. This data set is open source and collected by Kaur at al. (Kaur 2019).

The project includes Text Processing, Features Extraction, Machine Learning Modeling and Visualization of Results.

Text processing includes cleaning of data, removal of punctuation, stop words, numbers,

special characters, smileys, etc. Cleaning is to be done by Python libraries like pandas. Features extraction is done using different Tokenization and Vectorization methods like Bag of Words, TF-IDF Vectorizer, Count Vectorizer, etc.

Feature engineering will be conducted using different vectorizers techniques and explore the potential of Transfer Learning.

Different Machine Learning algorithms both Parametric and Non-Parametric will be applied after feature extraction where features and labels are inputs.

Labels include different categories of comments like gratitude, praising, suggestions, queries, etc. Based on the cross-validation and evaluation methods the final training and testing will be conducted for the whole data which gives the final results.

All the results will be represented using different Visualization techniques for a better understanding of models.

The major focus of this project is studying the effect different features extraction methods (vectorization) on Mix code language. And how these features extraction method effect the accuracy, precision of the models. In this project, the concept of Transfer Learning will be explored.

Kaur G, Kaushik A, Sharma S. Cooking is creating emotion: A study on hinglish sentiments of youtube cookery channels using semi-supervised approach. Big Data and Cognitive Computing. 2019 Sep;3(3):37.

### **b) Research objectives (Maximum 150 words)**

Please summarise briefly the objectives of the research.

- To investigate multiple Tokenization and Vectorization techniques on the text before Modelling. Tokenization is splitting of strings, sentences or documents into small chunks like words based on delimiter to understand text context. Vectorization is a process to convert text data into numerical vector forms which is useful for extracting features.
- To understand the feature engineering methods in Natural Language Processing (NLP). There will be many features extracted for text data in NLP. So, understanding of data is needed for best results.
- To examine the performance of models based on different evaluation methods like Accuracy, Precision, Confusion Matrices, Mean Square Errors, R2 Scores, etc.
- To observe the potential of different algorithms through both parametric and non-parametric models. Based on the cross-validation scores and evaluation methods on different models, Algorithm is selected for final Training and Testing.

### **c) Research location and duration**

<b>Location(s)/Population*</b>	Dundalk Institute of Technology, Dundalk, Ireland.
<b>Research start date</b>	01 <sup>st</sup> June 2022
<b>Research end date</b>	16 <sup>th</sup> September 2022

<b>Approximate duration</b>	4 Months
-----------------------------	----------

\* If location/Population other than DkIT campus/population, provide details of the approval to gain access to that location/population as an appendix.

### 3.3 Participants

		YES	NO	N/A
<b>Do participants fall into any of the following special groups?</b>	<b>Minors (under 18 years of age)</b>			N/A
	<b>People with learning or communication difficulties</b>			N/A
	<b>Patients</b>			N/A
	<b>People in custody</b>			N/A
	<b>People engaged in illegal activities (e.g., drug-taking)</b>			N/A
<b>Have you given due consideration to the need for satisfactory Garda clearance?</b>				N/A

### 3.4 Sample Details

<b>Approximate number</b>	N/A
<b>Where will participants be recruited from?</b>	N/A
<b>Inclusion Criteria</b>	N/A
<b>Exclusion Criteria</b>	N/A
<b>Will participants be remunerated, and if so in what form?</b>	
N/A	

<b>Justification for proposed sample size and for selecting a specific gender, age, or any other group if this is done in your research.</b>
N/A

### 3.5 Risk to Participants

- a) Please describe any risks to participants that may arise due to the research. Such risks could include physical stress, emotional distress, perceived coercion e.g., lecturer interviewing own students. Detail the measures and considerations you have put in place to minimize these risks
- b) What will you communicate to participants about any identified risks? Will any information be withheld from them about the research purpose or procedure? If so, please justify this decision.

### 3.6 Informed Consent

	YES	NO	N/A
Will you obtain active consent for participation?			N/A
Will you describe the main experimental procedures to participants in advance?			N/A
Will you inform the participants that their participation is voluntary and may be withdrawn at any point?			N/A
If the research is observational, will you ask for their consent to being observed?			N/A
With questionnaires, will you give participants the option of omitting questions they do not want to answer?			N/A
Will you tell participants that their data will be treated with full confidentiality and that, if published, it will not be identifiable as theirs?			N/A
Will the data be anonymous?			N/A
Will you debrief participants at the end of their participation?			N/A
Will your project involve deliberately misleading participants in any way, or will information be withheld?  If you answer yes, give details and justification for doing this below.			N/A

- a) **Please outline your approach to ensuring the confidentiality of data (that is, that the data will only be accessible to agree upon parties and the safeguarding mechanisms you will put in place to achieve this.) You should include details on how and where the data will be stored, and who will have access to it.**

The data is taken from

<https://archive.ics.uci.edu/ml/datasets/Youtube+cooking+channels+viewers+comments+in+Hinglish>.

No human information or personal data or survey data collected for this project. Anyone can download the dataset and I am citing the author of dataset in my report. No confidentiality is applicable for this data.

- b) **Please outline how long the data will be retained for, if it will be destroyed and how it will be destroyed.**

Standard DKIT procedures and guidelines will be followed where needed.

**1.Storage**

**2. Access**

**3. Communication**

**4. Digital Platform Usage**


**5. Data maintenance**



## Section 4

**Researcher** I have read, and I understand the DkIT Ethics Policy available from:

<https://www.dkit.ie/assets/uploads/documents/Research/Policies/DkIT%20Research%20Ethics%20Policy.pdf>

Signed:  Print Name: Murthy S Routhula Date: 25<sup>th</sup> March 2022  
(Researcher)

**Supervisor:** Applications for Ethical Approval of Undergraduate projects are forwarded to the Departmental Research Ethics Committee for approval or referral to the School Research Ethics Committee. Applications for Ethical Approval of Postgraduate and Staff projects are sent to the School Research Ethics Committee for Approval.

I have read and approved this form & information:

Signed: \_\_\_\_\_ Print Name: Dr. Abhishek Kaushik Date: \_\_\_\_\_  
(Supervisor/Head of Department/ Research Centre Director/ Head of School)

There is an obligation on the researcher and/or supervisor to bring to the attention of the Departmental/School Research Ethics Committee(s): (a) Any issues with ethical implications not clearly covered by this form (b) Any ethical issues which may arise during the carrying out of the research; (c) Any ethically significant change made to the project after approval.

## Section 5 (For office use only)

---

### STATEMENT OF ETHICAL APPROVAL (FOR UNDERGRADUATE PROJECTS ONLY)

This project has been considered using agreed department procedures and is now:

Approved:

☐

Referred to the School Ethics Committee:

☐

Signed: \_\_\_\_\_ Print Name: \_\_\_\_\_ Date: \_\_\_\_\_  
(Chair of Departmental Research Ethics Committee/Head of Department)

---

### STATEMENT OF ETHICAL APPROVAL

This project has been considered using agreed School procedures and is now:

Approved:

☐

Rejected (further information sought):

☐

#### Chair of School Research Ethics Committee

This project has been considered by the Ethics Committee and ethical approval is granted.

Signed: \_\_\_\_\_ Print Name: \_\_\_\_\_ Date: \_\_\_\_\_  
Chair of School Research Ethics Committee