Sri Lanka Institute of Information Technology

PROJECT REGISTRATION FORM

(This form should be completed and uploaded to the Cloud space on or before XXXXXXXXX)

The purpose of this form is to allow final year students of the B.Sc. (Hon) degree program to enlist in the final year project group. Enlisting in a project entails specifying the project title and the details of four members in the group, the internal supervisor (compulsory), external supervisor (may be from the industry) and indicating a brief description of the project. The description of the project entered on this form will not be considered as the formal project proposal. It should however indicate the scope of the project and provide the main potential outcome.

PROJECT TITLE (As per the accepted topic assessment form)	MAIML: Model Agnostic Interpretable Machine Learning
RESEARCH GROUP **(as per the Topic assessment Form)**	Data Science
PROJECT NUMBER	TMP-21-089 (will be assigned by the lecture in charge)

PROJECT GROUP MEMBER DETAILS: (Please start with group leader's details)

	STUDENT NAME	STUDENT NO.	CONTACT NO.	EMAIL ADDRESS
1	Kenneth Chamara J.K	IT18134704	0711537119	it18134704@my.sliit.lk
2	Abeyagunasekara S.H.P	IT18059564	0767866042	it18059564@my.sliit.lk
3	Perera G.Y.N	IT18113600	0770098395	it18113600@my.sliit.lk
4	Udari Kaushalya G.L	IT18056976	0763220555	it18056976@my.sliit.lk

SUPERVISOR, CO_ SUPERVISOR Details

SUPERVISOR Name	CO-SUPERVISOR Name
Mr. Prasanna Sumathipala	**Mr. Oshadha Seneweera**
Signature	**Signature**
Appendix 1	**Appendix 2**
23/01/2021	**23/01/2021**

EXTERNAL SUPERVISOR Details (if any, may be from the industry)

				Attach the email as Appendix 3
Name	Affiliation	Contact Address	Contact Numbers	Signature/Date

ACCEPTANCE BY CDAP MEMBER (This part will be filled by the RP team)

Name	Signature	Date

PROJECT DETAILS

Brief Description of your Research Problem: (extract from the topic assessment form)
When it comes to the Artificial Intelligence technology, Explainable AI(XAI) enables the ability to understand the results produced by the algorithms to the humans. This topic is highlighted with the "Black Box" machine learning models which cannot be explained why an algorithm came up with specific decision. Under the theory of Feature Significance, different methods are often applied through various research projects in Explainable AI area. Feature Importance-based methods define the result as what has been done, but do not include a solution to improve the outcome. For example, if a person applies for a bank loan and the system refuses the loan, the feature relevance method (such as SHAP, LIME) would provide what characteristics affect the outcome (salary) but not what could have been done to boost the outcome (increasing salary by ten thousand) It is particularly important to understand the domain-specific reasons behind the decisions made by the algorithm, not only for sensitive fields but also for all fields, as this will strengthen input factors that can affect the outcome of the Machine Learning Model. This research focuses on the implementation of a web-based application that utilizes a generic counterfactual interpretation system. The architecture is developed using algorithms of counterfactual generation and applied in different domains to different models of Black-Box Machine Learning. When applying training data, various types of machine learning algorithms function differently and the tuning of hyperparameters is unique to the models. These model categories must be considered when designing the structure and the framework should be implemented to be consistent with those models (Neural Networks, Decision Trees, Nearest Neighbour, SVM, linear models)
Description of the Solution: (extract from the topic assessment form) This research focuses on the implementation of a web-based application that utilizes a generic counterfactual interpretation system. The architecture is developed using algorithms of counterfactual generation and applied in different domains to different models of Black-Box Machine Learning. When applying training data, various types of machine learning algorithms function differently and the tuning of hyperparameters is unique to the models. These model categories must be considered when designing the structure and the framework should be implemented to be consistent with those models (Neural Networks, Decision Trees, Nearest Neighbour, SVM, linear models)

Main expected outcomes of the project: (extract from the topic assessment form)

Main Outcome:
Creation of a generalized counterfactual framework to boost the interpretability of black box models, allowing them to work on the feature importance as well.
Come up with an application that can be demonstrated the interpretability using counterfactual framework.

Outcome 1:
Implement the components and tools to explain a neural network using counterfactual explanation method.

Outcome 2:
The main disadvantage of the TreeSHAP is TreeSHAP can produce unintuitive feature attributions. Develop a method to eliminate or minimize the unintuitive feature attributions from the TreeSHAP using counterfactual methodologies.

Outcome 3:
Using a counterfactual algorithm, generate a Support Vector Machine model and optimized for a better solution, and implementing necessary components to utilize the SVM model related to the framework.

Outcome 4:
Generate and optimize a K-Nearest Neighbour model using a counterfactual algorithm for a better solution and implement the necessary components to utilize the KNN model related framework.

WORKLOAD ALLOCATION (**extract from the topic assessment form after correcting the suggestions given by the topic assessment panel.**)
(Please provide a brief description about the workload allocation)

MEMBER 1

- Implementing related components to apply a Neural Network to the framework.
- Implement local rule extraction.
- Creating debugging tools to test and identify counterfactuals generated for neural network.
- Training a neural network and explaining using the framework

MEMBER 2

- Creating a component to apply counterfactual algorithm which can used for Black-Box model – Boosted Tree.
- Develop an explainable boosted tree algorithm which utilize the counterfactual methods.
- Generate graphs to determine how each feature value effects the change in output for decision trees.
- By using graphs generated use a counterfactual algorithm to determine which feature affects the outcome of the trees most.
- Using this approach determine the feature importance and thereby find a method to identify unintuitive feature generation attributions generated by the TREEShap.

MEMBER 3

- Creating necessary components to apply Support Vector Machine model in the framework.
- Implement a method to figure out the feature dependency.
- Optimize Counterfactual algorithm for SVM model.

Training the model and provide the relevant explanation from the implemented framework**.**

MEMBER 4

- Implement components of to apply the K-Nearest Neighbour model to the framework.
- Planning and implementing virtualisation components to the framework
- Training a KNN model and explaining using the framework
- Optimize counterfactual algorithm to KNN .

DECLARATION (Students should add the Digital Signature)

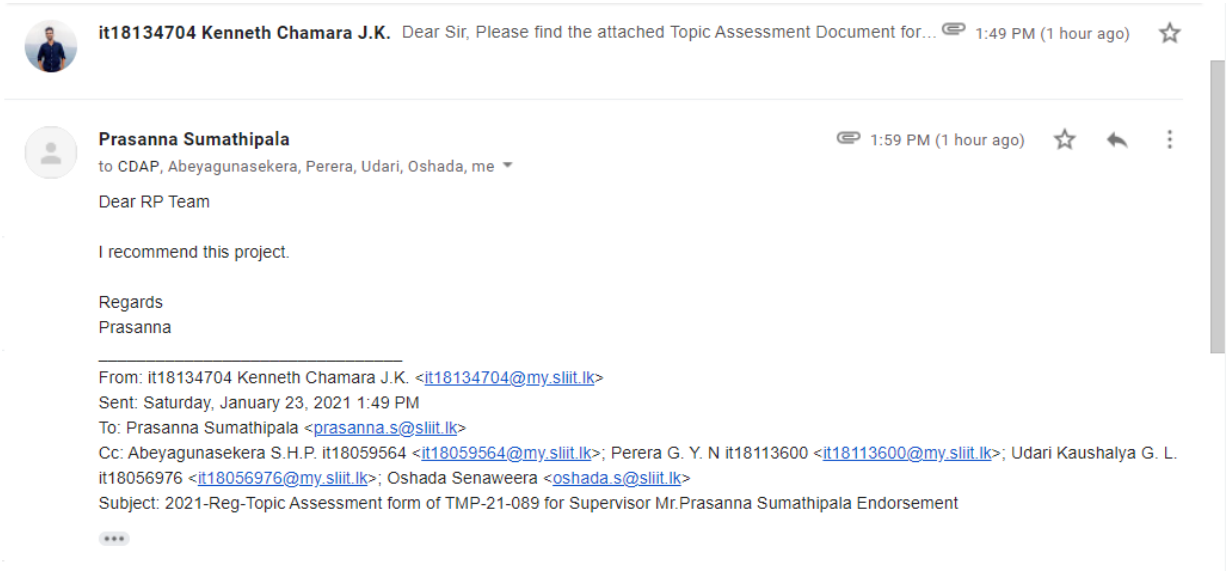
"We declare that the project would involve material prepared by the Group members and that it would not fully or partially incorporate any material prepared by other persons for a fee or free of charge or that it would include material previously submitted by a candidate for a Degree or Diploma in any other University or Institute of Higher Learning and that, to the best of our knowledge and belief, it would not incorporate any material previously published or written by another person in relation to another project except with prior written approval from the supervisor and/or the coordinator of such project

and that such unauthorized reproductions will construe offences punishable under the SLIIT Regulations.

We are aware, that if we are found guilty for the above mentioned offences or any project related plagiarism, the SLIIT has right to suspend the project at any time and or to suspend us from the examination and or from the Institution for minimum period of one year".

	STUDENT NAME	**STUDENT NO.**	**Signature**
1	Kenneth Chamara J.K	IT18134704	
2	Abeyagunasekara S.H.P	IT18059564	
3	Perera G.Y.N	IT18113600	
4	Udari Kaushalya G.L	IT18056976	

Appendix 1



it18134704 Kenneth Chamara J.K. Dear Sir, Please find the attached Topic Assessment Document for... 1:49 PM (1 hour ago)

Prasanna Sumathipala 1:59 PM (1 hour ago)
to CDAP, Abeyagunasekera, Perera, Udari, Oshada, me

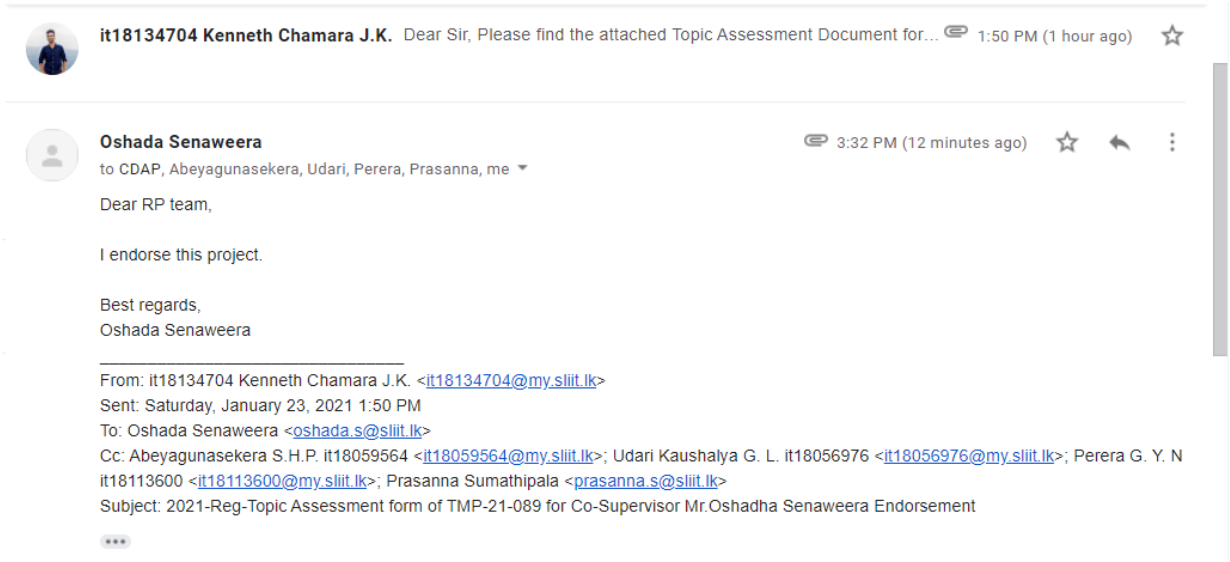
Dear RP Team

I recommend this project.

Regards
Prasanna

From: it18134704 Kenneth Chamara J.K. <it18134704@my.sliit.lk>
Sent: Saturday, January 23, 2021 1:49 PM
To: Prasanna Sumathipala <prasanna.s@sliit.lk>
Cc: Abeyagunasekera S.H.P. it18059564 <it18059564@my.sliit.lk>; Perera G. Y. N it18113600 <it18113600@my.sliit.lk>; Udari Kaushalya G. L. it18056976 <it18056976@my.sliit.lk>; Oshada Senaweera <oshada.s@sliit.lk>
Subject: 2021-Reg-Topic Assessment form of TMP-21-089 for Supervisor Mr.Prasanna Sumathipala Endorsement

Appendix 2



it18134704 Kenneth Chamara J.K. Dear Sir, Please find the attached Topic Assessment Document for... 1:50 PM (1 hour ago)

Oshada Senaweera 3:32 PM (12 minutes ago)
to CDAP, Abeyagunasekera, Udari, Perera, Prasanna, me

Dear RP team,

I endorse this project.

Best regards,
Oshada Senaweera

From: it18134704 Kenneth Chamara J.K. <it18134704@my.sliit.lk>
Sent: Saturday, January 23, 2021 1:50 PM
To: Oshada Senaweera <oshada.s@sliit.lk>
Cc: Abeyagunasekera S.H.P. it18059564 <it18059564@my.sliit.lk>; Udari Kaushalya G. L. it18056976 <it18056976@my.sliit.lk>; Perera G. Y. N it18113600 <it18113600@my.sliit.lk>; Prasanna Sumathipala <prasanna.s@sliit.lk>
Subject: 2021-Reg-Topic Assessment form of TMP-21-089 for Co-Supervisor Mr.Oshadha Senaweera Endorsement