IT4010 – Research Project - 2024 Topic Assessment Form

Project ID: **R24-078**

1. Topic (12 words max)

Air Quality Monitoring and Analysis Based Predictive System

2. Research group the project belongs to

Knowledge Inspired Computing (KIC)

3. Research area the project belongs to

ICT for Development (ICTD)

4. If a continuation of a previous project:

Project ID	R24-078
Year	2024

5. Brief description of the research problem including references (200 – 500 words max) – references not included in word count.

Air pollution, both indoor and outdoor has become a concerning issue worldwide. Among the air pollutants Particulate Matter (PM) Carbon Dioxide (CO2) Carbon Monoxide (CO) Ground level Ozone (O3) Nitrogen Oxides (NOx) Sulfur Oxides (SOx) and Lead (Pb) are the significant ones found in our environment. This environmental health problem affects people, in both developing countries.

When it comes to air pollution emissions from motor vehicles, fuel burning and industrial activities are the primary sources. In Sri Lanka 24,000 deaths were attributed to air pollution in 2012 [1]. The respiratory health symptoms resulting from air pollution include lower respiratory tract infections, pneumonia, bronchitis, asthma, reduced lung functions and growth as well, as lung cancer [2]. In 2016 ambient (outdoor) air pollution was estimated to cause around 4.2 million deaths globally in both urban and rural areas [3].

The primary source of air pollution, in Sri Lanka is the emissions from vehicles, which account for than 60% of the emissions in Colombo [4]. Considering the population in Sri Lanka school children are particularly vulnerable to levels of pollutants due to overcrowding in schools located in cities, especially Colombo and Kandy. The levels of air pollutants were significantly greater, within the premises of schools when compared to schools [5].

Sri Lanka is facing a serious air pollution problem that severely impacts the daily life of every Sri Lankan. The main source of ambient air pollution in Sri Lanka is vehicular emissions. A methodology to monitor the air quality in real-time with an overall coverage of Sri Lanka, and automatically process these huge data to identify air quality levels in a **specific area** is now becoming a timely research topic.

An air quality monitoring and analysis based predictive system is proposed to monitor the ambient air quality, provides the best route with minimum polluted air, maps the heatmaps to identify the current air quality of an area easily and predict the future air quality of each area. The prototype was implemented by hierarchically deploying two different gas sensors, an Arduino Uno board and a Wi-Fi module, to implement in open spaces between smart buildings, and transfers the sensor data back to the information processing center by using IoT technology for real-time display. The information processing center stores real-time information which is collected from the sensors to the database. By reading sensor data stored in the database, the front-end system draws real-time, accurate air quality levels included maps and predicts the less polluted routes and the air quality level over an area.

Reference:

[1] UNDP, "MAPS Approach Supporting SDG Implementation in Sri Lanka," 2018.

[2] HEI International Scientific Oversight Committee, "Outdoor air pollution and health in the developing countries of Asia: a comprehensive review. Special report 18," Heal. Eff. Institute, Bost., no. November 2010.

[3]WHO, "Ambient (outdoor) air quality and health," 2018. [Online]. Available: https://www.who.int/news-room/factsheets/detail/ambient-(outdoor)-air-quality-and-health. [Accessed: 23-Apr-2019].

[4]J. B. Batagoda B, Sugathapala A, Yalegama M, "Urban Air Quality Management in Sri Lanka. Colombo: Air Resource Management Center (AirMAC)," 2004.

[5] Y. L. S. Nandasena, A. R. Wickremasinghe, and N. Sathiakumar, "Air pollution and health in Sri Lanka: A review of epidemiologic studies," BMC Public Health, vol. 10, 2010.

6. Brief description of the nature of the solution including a conceptual diagram (250 words max)

To address this problem, An innovative solution proposed for the above scenario would be a prototype which uses an air quality monitoring device and a mobile application to visualize air quality information in terms of route suggestions, heat map, and air quality predictions. People’s awareness is the main objective of this research, and it helps people who are suffering from respiratory diseases, kids and older persons to avoid the most polluted areas by identifying highly polluted areas when traveling.

Pollution Levels Prediction:

The primary objective of this component is to develop a robust model for predicting pollution levels. This involves analyzing historical data, current environmental conditions, and other relevant factors to forecast the concentration of pollutants in the air accurately. Through rigorous analysis, the project team pinpoints the variables with the most substantial impact. Subsequently, feature engineering techniques are applied to enhance the model's predictive capabilities. This involves transforming and manipulating the chosen features to extract more valuable information and improve the overall accuracy of the predictive model. The predictive model introduces dynamic predictions, a unique feature that enables real-time adjustments based on changing environmental factors. This innovation enhances the model's responsiveness to rapidly evolving conditions. Furthermore, the integration with the route generation component ensures that the pollution prediction model actively contributes to the generation of optimal and least-polluted paths.

To further strengthen the system's transparency and user awareness, a **comparison** with real-time pollution data is implemented. This comparison serves not only for accuracy validation but also to showcase potential conflicts or deviations. A **graphical representation** is employed to illustrate the comparison between past, present, and future predicted data. The graph provides a visual narrative, allowing users to discern patterns, anomalies, and potential discrepancies between real-time observations and model predictions. This comprehensive approach enhances the system's credibility, providing users with a holistic understanding of the environmental dynamics and fostering greater trust in the predictive capabilities of the air quality monitoring system.

Sensor based device creation and value accuracy:

In this component, MQ4, and MQ7 sensors were used to measure Methane (CH_4) and Carbon Monoxide (CO) gases in the air respectively. The conductivity increases with the increment of gas concentration in the polluted air, when polluted air goes through these sensors.

The difference of conductivity can be converted into gas concentration values, by implementing an Arduino code using the Arduino Uno board. The output analog value (0 - 4095) of each sensor should be converted into sensor voltage value using the (1).

Sensor Voltage = (Analog Reading * 3.3V) / 4095 (1)

After sensor voltage calculation, we can convert that value into a Parts per Million (PPM) concentration value using the sensitivity calibration curve of each gas sensor given in their datasheets separately, which were drawn under the standards between the two axis; the gas concentration (PPM) and sensor voltage (V). With the aid of those sensitivity curves of each gas sensor, built two relationships as (2) and (3) for the sensors separately, to convert the sensor voltages into concentration value. Using the graphical analysis software, Engauge Digitizer; points of these graphs were selected and drew an exponential to that graph using the selected points. With the aid of that drawn exponential, declared equations to each graph for building the relationship between sensor voltage and gas concentration in ppm. The equations built for the relationship of gas concentration (ppm) and sensor voltage (V_RL) of two sensors are as follows.

CH4 concentration = 10.938 * e(1.7742*V_RL) (2)

CO concentration = 3.027 * e(1.0698*V_RL) (3)

The ESP8266 Wi-Fi module is used to transmit the sensor data to the database in every one minute of time interval to make the device more real in time. The location of the device is identified using a predefined terminal number to the places where the devices implemented. Instead of using a GPS system, this predefined terminal number system was used for reducing the size of the data packet, as this system is a more real-time system, and it uploads the data in every 1 minute of the time interval. This allocated terminal number is also uploaded with the sensor data for the ease of identification of device location.

The data packet is sent by using a Hypertext Transfer Protocol (HTTP) Post request. The packet uses encoded format when sending to the server, to increase the security of the data packet. A Transmission Control Protocol (TCP) connection is established with the server by the Wi-Fi module. An Arduino Uno board is used to implement the data processing code. This Arduino code handles the sensors to get the sensor inputs, process them and handles the Wi-Fi module to send these processed outputs to the database.

In addition to the aforementioned components, the designed hardware **enhances user notification capabilities.** Specifically, the hardware is equipped to send email notifications to registered email addresses whenever the sensor detects high readings for particulate matter or Carbon monoxide. This notification feature adds an extra layer of proactive communication, alerting users to potential environmental risks and

Real-time pollution map:

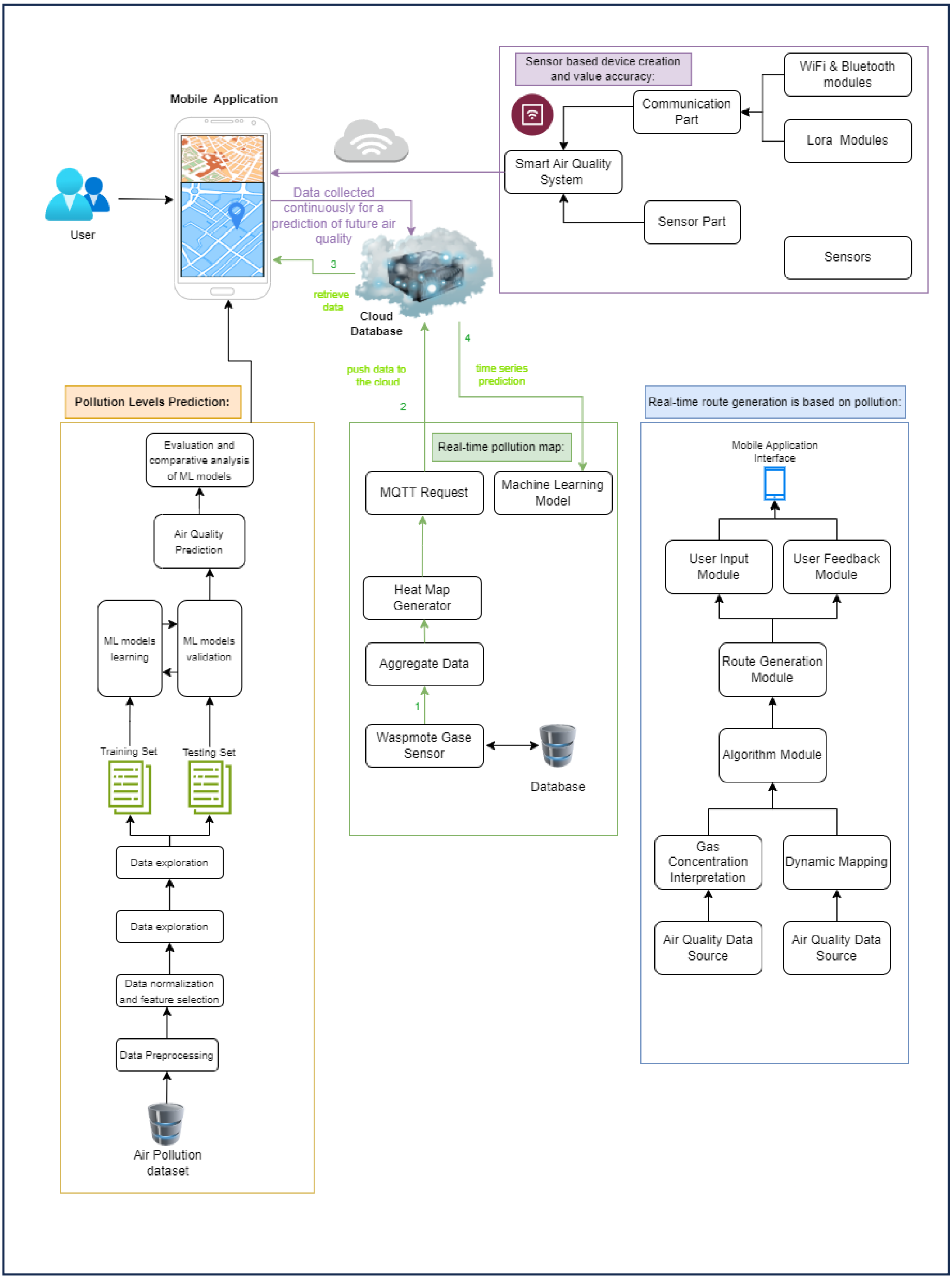
This component focuses on providing a **visual representation of real-time pollution levels** across different regions to facilitate informed decision-making. It is a graphical representation of data that uses a system of color-coding to represent different values. It used in various forms of analytics but is most commonly used to show user behavior on specific situation. There are two heat maps were generated for methane and carbon monoxide separately. It is generated using accurate sensor data stored in the database, JavaScript and HTML language. Latitude and longitude are related to the terminal number, which is used to identify the location. Then using color schemes, the sensor values related to the location are displayed separately in two maps. Moreover, it displays the real time level of air pollution in an area. Many different color schemes can be used to illustrate the heat maps. In the mobile application, users can use the heat maps to identify the air polluted in an area, due to high amounts of various gases in that area.

In addition to its primary function of providing a visual representation of real-time pollution levels across different regions, this component incorporates cutting-edge technology to enhance user experience. Leveraging Natural Language Processing (NLP), a voice assistant has been integrated, allowing users to obtain information about areas with high pollution levels through **spoken queries**. Furthermore, language accessibility is prioritized with the implementation of a **translator** supporting both Tamil and English languages. This innovative approach not only empowers users to visually assess pollution data but also enables seamless interaction and understanding of the information using advanced language technologies.

Real-time route generation is based on pollution:

The aim of the sophisticated "Real-time route generation based on pollution" component is to provide users with the most optimal routes to their destinations. This cutting-edge system takes into consideration both the distance and air quality factors. To achieve this, an innovative algorithm, which combines Dijkstra's algorithm with a dynamic mapping process, is employed. By modifying Dijkstra's algorithm to include pollution levels, the system ensures that users commence their journeys from the closest non-polluted point, thereby minimizing their exposure to unhealthy air quality. In this process, gas concentration data, measured in Parts Per Million (PPM), plays a critical role, with color-coded interpretations indicating the severity of pollution.

The algorithm analyzes real-time mapping data and gas concentrations to suggest routes that not only minimize distance but also prioritize areas with cleaner air. These suggested routes, which incorporate visual markers, such as red for starting points, blue for endpoints, and green for non-polluted areas, are then seamlessly integrated into a mobile application. This user-friendly interface enables individuals to input their desired destinations and receive visually represented routes that prioritize their health and safety. The solution is designed for real-life applications, providing users with a comprehensive travel experience that not only emphasizes efficiency but also takes into account the well-being of individuals by avoiding areas with high pollution levels.


Mobile Application
User
Data collected continuously for a prediction of future air quality
3
retrieve data
Cloud Database
push data to the cloud
2
4
time series prediction
Sensor based device creation and value accuracy:
Communication Part
WiFi & Bluetooth modules
Lora Modules
Smart Air Quality System
Sensor Part
Sensors
Pollution Levels Prediction:
Evaluation and comparative analysis of ML models
Air Quality Prediction
ML models learning
ML models validation
Training Set
Testing Set
Data exploration
Data exploration
Data normalization and feature selection
Data Preprocessing
Air Pollution dataset
Real-time pollution map:
MQTT Request
Machine Learning Model
Heat Map Generator
Aggregate Data
1
Waspmote Gase Sensor
Database
Real-time route generation is based on pollution:
Mobile Application Interface
User Input Module
User Feedback Module
Route Generation Module
Algorithm Module
Gas Concentration Interpretation
Dynamic Mapping
Air Quality Data Source
Air Quality Data Source

7. Brief description of specialized domain expertise, knowledge, and data requirements (300 words max)

The proposed solution requires expertise in fields, including science, data science, hardware development and mobile application programming. Firstly a deep understanding of science is crucial to develop a pollution prediction model. This involves analyzing data studying conditions and identifying the factors that impact air quality. Data scientists play a role, in enhancing the accuracy of the model by utilizing feature engineering techniques and making real time adjustments based on changing factors.

Secondly expertise in hardware development is essential for designing an air quality monitoring device. Knowledge of sensors like MQ4 and MQ7 is necessary to measure Methane and Carbon Monoxide gases present in the air. The creation of a hardware system involves coding with Arduino and calibrating curves to ensure that sensor readings are converted into Parts per Million (PPM) concentration values. Additionally familiarity with communication protocols, like HTTP and TCP is required to transmit sensor data to a database using the ESP8266 Wi Fi module.

Furthermore, it is essential to have an understanding of mobile app development in order to create a user interface that effectively presents air quality data. Proficiency, in JavaScript, HTML and mobile app programming is vital for generating pollution maps. Additionally advanced language technology skills are necessary, for incorporating a voice assistant using Natural Language Processing (NLP).

Moreover, domain expertise in route optimization algorithms is essential for developing the real-time route generation component. Combining Dijkstra's algorithm with dynamic mapping processes requires proficiency in algorithm development and GIS (Geographic Information System). The interpretation of gas concentration levels and the corresponding color coding for risk assessment demands knowledge of environmental health and safety standards.

In terms of data requirements, access to historical and real-time environmental data, including pollution levels and meteorological conditions, is essential for training and validating the pollution prediction model. Additionally, a database infrastructure capable of storing and managing sensor data is required. The system relies on accurate sensor readings and geographical information for real-time pollution map generation and route optimization, necessitating a robust data collection and storage mechanism.

In summary, the proposed solution demands a comprehensive skill set spanning environmental science, data science, hardware development, mobile application programming, and GIS. Access to accurate and diverse datasets is critical for developing, testing, and improving the various components of the innovative air quality monitoring system.

8. Objectives and Novelty

Main Objective: Develop a prototype which uses an air quality monitoring device and a mobile application to visualize air quality information in terms of route suggestions, heat map, and air quality predictions. People's awareness is the main objective of this research, and it helps people who are suffering from respiratory diseases, kids and older persons to avoid the most polluted areas by identifying highly polluted areas when traveling.			
Member Name	Sub Objective	Tasks	Novelty
Inthikhaff M. I. M (IT21058028) Group Leader	Develop a Robust Pollution Prediction Model	**Task 1:** Collecting polluted air data from the emission standard publication websites and Preprocessing, Cleaning, Transforming. **Task 2:** Feature Selection and Analysis. **Task 3:** Model Development Implement machine learning algorithms for pollution level prediction. Explore and select the most suitable predictive model based on performance metrics.	Introduction of dynamic predictions to adapt to real-time changes, enhancing the model's responsiveness. The integration of rigorous analysis, variable impact assessment, and feature engineering contributes to the development of a robust and accurate pollution prediction model. The introduction of dynamic predictions ensures the model adapts to real-time changes, improving its accuracy and effectiveness.

		Task 4: Dynamic Predictions for Real-time Adjustments **Task 5:** Enhanced User Notification Capabilities **Task 6:** Real-time Pollution Data Comparison and Visualization	Equipping the hardware with email notification capabilities adds an extra layer of proactive communication, enhancing user awareness and promoting informed decision-making. The comprehensive graphical representation fosters transparency, allowing users to understand the environmental dynamics and build trust in the predictive capabilities of the air quality monitoring system.
Rimas M. J. M (IT21003714)	Sensor based device creation and value accuracy.	**Task 1:** Sensor Calibration **Task 2:** Arduino Code Implementation **Task 3:** Relationship Building for Gas Concentration Calculation. **Task 4:** Real-time Data Transmission and Location Identification. **Task 5:** Secure and Efficient Data Transmission Arduino Code for Data Processing and Transmission	The calibration process and Arduino code implementation ensure accurate conversion of sensor readings to meaningful gas concentration values. Creating exponential equations from sensitivity curves enhances accuracy and precision in converting sensor voltage to gas concentration, ensuring robust data reliability.

			Using a predefined terminal number system for location identification reduces data packet size and ensures real-time data transmission. Implementing secure HTTP Post requests and TCP connections enhances data security and efficiency during transmission. The Arduino code acts as a central processing unit, ensuring accurate sensor data processing and secure transmission to the database.
Kumari J. M. D (IT21152832)	Real-time Pollution Visualization and Advanced Interaction	**Task 1:** Heat Map Generation for Methane and Carbon Monoxide. **Task 2:** Mobile Application Integration and User Interaction **Task 3:** Integration of Voice Assistant through NLP	Real-time, location-based heat maps provide users with an immediate and accurate visual representation of pollution levels. Mobile application integration enhances accessibility, allowing users to conveniently assess pollution levels on-the-go.

			Incorporating a voice assistant adds an interactive layer, enabling users to obtain pollution information effortlessly through spoken queries. Language accessibility through translation support ensures a wider user base, allowing users to interact with the system in their preferred language. dynamic time-dependent maps
Pasan M.G. R(IT21057588)	Real-time Route Generation Based on Pollution	**Task 1:** Development of Pollution-Aware Route Generation Algorithm. **Task 2:** Risk-based Start and End Point Selection **Task 3:** Mobile Application Integration and Map Display	The novel approach involves dynamically mapping the optimal route based on real-time pollution data, ensuring user safety and health during transit. Selecting routes based on the nearest non-polluted points and implementing

		Task 4: Route Testing and Validation	color-coded markers enhance user understanding and safety. Real-time route suggestions based on pollution levels are visually represented on a map, providing users with actionable information. Continuous testing and optimization ensure the reliability and responsiveness of the route generation system under diverse conditions.

9. Supervisor checklist

a) Does the chosen research topic possess a comprehensive scope suitable for a final-year project?

Yes	√	No	

b) Does the proposed topic exhibit novelty?

Yes	√	No	

c) Do you believe they have the capability to successfully execute the proposed project?

Yes	√	No	

d) Do the proposed sub-objectives reflect the students' areas of specialization?

Yes	√	No	

e) Supervisor's Evaluation and Recommendation for the Research topic:

Students are required to create a device to capture primary data.

Device validation and accuracy is a major requirement in the project.

10. Supervisor details

	Title	First Name	Last Name	Signature
Supervisor	Ms.	Chathurangika	Kahandawaarachchi	Chathurangika.K (Signed)
Co-Supervisor				
External Supervisor				
Summary of external supervisor's (if any) experience and expertise				

IT4010 – Research Project - 2024 Topic Assessment Form

This part is to be filled by the Topic Screening Panel members.

Acceptable: Mark/Select as necessary

Topic Assessment Accepted	
Topic Assessment Accepted with minor changes (should be followed up by the supervisor)*	
Topic Assessment to be Resubmitted with major changes*	
Topic Assessment Rejected. Topic must be changed	

* Detailed comments given below

Comments

The Review Panel Details

Member's Name	Signature

*__Important__:

1. According to the comments given by the panel, make the necessary modifications and get the approval by the **Supervisor** or the **Same Panel**.
2. If the project topic is rejected, identify a new topic, and request the RP Team for a new topic assessment.
3. The form approved by the panel must be attached to the **Project Charter Form**.