|  |  |
| --- | --- |
| **Student:** |  |
| **Student ID and email:** |  |
| **Course:** |  |
| **Supervisor:** |  |
| **Second Reader**: |  |

# CIS4055: Computing Masters Project Proposal

## Project Title

*Lung Cancer Risk Prediction Using Machine Learning and Web-Based Diagnostic Tool*

## What problem are you trying to solve?

Lung cancer is one of the leading causes of cancer-related deaths worldwide, and its early detection is critical for improving survival rates. However, many individuals do not recognize the symptoms early enough, and access to timely medical screening can be limited. The goal of this project is to develop a machine learning-based system that can predict the risk of lung cancer based on user-inputted health data and symptoms. By integrating this model into a user-friendly web application, the system aims to provide an accessible, early-warning tool that helps individuals assess their potential risk and seek medical advice sooner.

## Research Question.

Can machine learning algorithms accurately predict an individual's risk of lung cancer based on health-related input data, and can this prediction be effectively delivered through a user-friendly web application to support early detection and awareness?.

## Hypothesis.

The use of machine learning algorithms can significantly improve the accuracy of predicting lung cancer risk based on individual health data and symptoms, and integrating the best-performing model into a web application will provide a practical and accessible tool for early risk assessment.

## Objectives of the project.

1. To investigate previous research and existing systems related to lung cancer prediction using machine learning techniques.

2. To collect, preprocess, and explore lung cancer-related datasets through Exploratory Data Analysis (EDA).

3. To apply and compare multiple machine learning algorithms to identify the most accurate model for lung cancer risk prediction.

4. To design and implement a predictive model based on the best-performing algorithm.

5.To develop a user-friendly web application that allows individuals to input personal data and receive lung cancer risk predictions.

6. To critically evaluate the performance and usability of the final system through accuracy metrics and user feedback.

## Expected end products.

*The primary deliverable will be a detailed dissertation report that documents the entire research and development process, including problem definition, literature review, data analysis, model development, and evaluation.*

*In addition to the report, I will develop the following artefacts:*

* *A fully functional web-based application that allows users to input their health-related data and receive a prediction regarding their potential risk of lung cancer.*
* *Results of technical experiments comparing multiple machine learning algorithms for accuracy, precision, recall, and ROC curve performance.*
* *Evaluation and justification for the chosen algorithm based on experimental results.*
* *Application of data science techniques, including Exploratory Data Analysis (EDA), feature correlation analysis, model training and testing, and performance evaluation on a real-world health dataset.*
* *Insights and guidance on using machine learning for early disease detection, demonstrating how such technologies can support public health initiatives through accessible digital tools.*

## Ethical Considerations.

< Please review the guidance on the Bb module for more information about this, but here you should outline what ethical, legal and professional issues you will need to address in your project. You should also consider if your project will involve any human input e.g.: survey participants, testers/evaluators of your product or artefact? How will you protect their privacy? How will you ensure the security of any data that results from this? Is the content matter of your product/artefact likely to put someone, including yourself, at risk of harm (physical or psychological)?

As discussed on the Bb module you need to determine if your project is low risk or whether you need to apply for Ethical approval via the ERM or not. Please complete the table below and include it in your report! >

### Ethics Declaration:

|  |  |  |
| --- | --- | --- |
|  | **True** | **False** |
| My project is entirely literature based and/or technical, |  |  |
| My Project does not use any form of participants, |  |  |
| My Project does not use external inputs (e.g. liaising with someone in industry), |  |  |
| My Project does not require me to do work off campus (e.g. in a company), |  |  |
| My Project does not use secondary data sets |  |  |

**All the students need to apply for ethics clearance through ERM (see Blackboard for more information). If you answer “False” to any of these statements, then it will go to Ethics review committee otherwise your supervisor will receive an email to approve**.

## Justification for Masters level project.

This project is appropriate for a Master’s level dissertation because it combines advanced machine learning techniques with practical software development to address a real-world health problem—early detection of lung cancer. It demonstrates the application of multidisciplinary knowledge across data science, artificial intelligence, and web development.

The project involves:

* **Technical Complexity**: Comparative analysis of multiple machine learning algorithms, hyperparameter tuning, performance evaluation using advanced metrics (ROC, AUC, confusion matrix), and selection of the most effective model.
* **Applied Research**: Although secondary data is used, the project performs original experimentation and model development, generating new insights from the data.
* **Novel Application**: The integration of a predictive model into a web-based application for public use reflects a novel and practical implementation, not just theoretical analysis.
* **Critical Thinking and Evaluation**: The project requires evaluating model effectiveness, data limitations, and ethical implications, showcasing the depth expected at the Master’s level.
* **Professional Skills**: Building a secure, user-friendly application aligns with industry practices and demonstrates real-world relevance and applicability.

Together, these elements ensure the project goes beyond undergraduate-level work and meets the academic and professional expectations of postgraduate study.