Student Research Plan Template

Table of Contents

PROJECT INFORMATION

Student Name(s): Samarth Singhal, Nihal Takalkar, Swar Patel

Student School: Walsh Middle School

Title of Project: Lung Cancer

Category: TME

Team Project (yes or no): Yes

Date you plan to start: 11/26

AlWAYS USE COMPLETE SENTENCES AND ANSWER THE FOLLOWING

Research Location: *(Where do you plan to do your research check all that apply)*

☐Home

☐School

☐Regulated Research Institution (i.e. University Lab etc.)

☐Place of Business, Commercial Property or Industrial Setting

☐Field or Other Public Location (i.e. City Park, Lake, River, Shopping Mall etc.)

☐Other (Neighbor's house, mentor's house etc.)

Location Address: *(For each location above, list the name and give the street address if available, or if not give location description or GPS coordinates)*

4309 Pasada Ln

James Garland Walsh

Rationale: *(Why is this project important? Include a brief summary of the background that supports your research problem and explain why this research is important scientifically and if applicable, explain any societal impact of your research.)*

Our project is important because it can help save many lives by detecting lung cancer early.

Hypothesis/Engineering Goal: (*State your HYPOTHESIS(ES), RESEARCH QUESTION(S), ENGINEERING GOAL(S), EXPECTED OUTCOMES. How is this related to your reasons for choosing this project?)*

Our expected AI will be expected to predict lung cancer either using the symptoms that the user inputs or the patient can upload a ct scan of the chest and predict it with a high accuracy rate.

Materials: *(List the materials you will be using, including chemicals and amounts and concentrations. Include where you purchased any bacteria or other potentially hazardous biological agents. If you used E. coli k-12, make sure to list that as the material.)*

Laptops

Kaggle

Procedure: *(Describe in detail all procedures and experimental design including methods for data collection, and when applicable, the source of data used.* ***Describe only your project. Do not include work done by your mentor or others.*** *If you will use published surveys, questionnaires or tests, describe how you obtained these, including required permission if applicable.)*

Environment setup:

Step 1: Made a repository on GitHub call LUNG-CANCER-AI

Step 2: Then we downloaded Github desktop and git

Step 3: Using github we cloned the repository into a local directory called:
'C:\Users\samar\src\Lung-Cancer-AI'

Step 4: We downloaded and installed vscode

Step 5: We opened source directory specified in step 3 in vscode and updated this read me

Step 6: Installed python and said hello world for the millionth time

Step 7: Added a requrements.txt file for all the libraries i think i will need for now

Step 8: Installed the requirements using the following command 'pip install -r .\requirments.txt'

Step 8.1: Every time I need a new library i can add it in requirements and run the code above again

Data Acquisition:

Step 1: Downloaded Lung Cancer Dataset from Kaggle.com from 'https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer/'

Step 2: Copied this file into the data directory for this project

Step 3: renamed the file to 'survey_lung_cancer.csv'

Understanding the Data:

Step 1: File is a csv file(comma separated value)

Step 2: We will open the file using pandas and examine the contents

Step 3: I read the csv using the read_csv function in the pandas library

Step 4: I print the number of rows and columns using df.shape

Step 5: I print all the name of the columns using df.columns

Step 6: The data is of current patients who tested for lung cancer and were given a diagnosis of positive or negative in the column LUNG_CANCER. The other column represents various attributes that were collected for the patient such as age and smoking.

Step 7: Our task in this project is to use this data to predict whether or not someone has lung cancer if we are provided all the attributes in this file.

Splitting the Data:

1. I got all the data in a jupyter notebook and then i converted the dataframe into two dataframes X and y. X represents the predictors which in this case are the symptoms and y is the result which in this case is whether the patient has lung cancer or not

2. now we split the data into train and test datasets using the command 'X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)'

3. This splits 80 percent of data for training the ai and the remaining twenty for testing the ai model.

Choosing the Algorithm:

1. This problem requires us to use supervised learning to predict/classify if the patient has lung cancer or not

2. Now we go and search algorithms available for classification.

3. I tried to make the logistic regression due to logistic being yes or no and i fail on the second line of code

4. We took a little detour due to an error because the values for some columns were strings so we made them into numbers so that we could easily use logistic regression. *We tried to make logistical regression, is the reason it didn't work

5. Today I finished most of the AI. I just have to integrate it into a ai and then I have to finish the posterboard tomorrow.

Creating web UI for this project

1. Import flask in requirements.txt and install using pip

2. Made a basic web page that we will expand on

3. while expanding a note to self is that yf = yellow fingers.

4. We added all of the possible symptoms to the html

5. we made it to drop down using this'<label for="fatigue">Fatigue:</label>

<select id="fatigue" name="fatigue">

  <option value="Yes">Yes</option>

  <option value="No">No</option>

</select>
'

6. We made the lung cancer we just have to clean up the website

Risk and Safety: *(Identify any potential risks and safety precautions needed.)*

No risk at all

Data Analysis: *(Describe how you will analyze or compare your data to determine your results. (Will you use tables, graphs or formulas?)*

there is a function built into python pandas library that can calculate the accuracy for you

Bibliography:

Type in your Bibliography: *(Must have at least 5 valid sources, use APA format)*

Author(s) Title Source Type

GitHub Copilot GitHub Copilot https://github.com/copilot Website

OpenAI GPT-4.0 https://openai.com/blog/gpt-4/ Website

Sandeep Singhal

Mysar Ahmedabad Lung Cancer Detection Using Convolutional Neural Networks (CNNs) Kaggle Discussion Post on https://www.kaggle.com/competitions/siim-acr-pneumothorax-segmentation/discussion/322516 Discussion post

K Scott Mader Lung Cancer Detection Using Convolutional Neural Networks (CNNs) Kaggle Discussion Post on https://www.kaggle.com/competitions/siim-acr-pneumothorax-segmentation/discussion/322516 Discussion post

Extra Information:

Choose the appropriate sections needed for your project. Then fill out those sections on the following pages. Delete the sections you do not need.

- ☐ Human Participant Studies
- ☐ Vertebrate Animal Studies
- ☐ Potentially Hazardous Biological Agents
- ☐ Hazardous Materials/Activities/Devices
- ☐ Continuation Projects

HUMAN PARTICIPANT STUDIES (*complete if applicable, this includes studies where the researcher is a participant, delete if not needed*)

Participant Details: *(Describe who will participate in your study (age range, gender, racial/ethnic composition). Identify any vulnerable populations (minors, pregnant women, prisoners, mentally disabled or economically disadvantaged)*

Recruitment: *(Where will you find or get your participants?)*

Consent: *(You need to get permission from your participants and inform them of your study, will it be Verbal or Written?)*

Activities: *(What will participants be asked to do?)*

Time: *(How long will it take for each participant to participate?)*

Surveys: *(Will you be handing out any surveys, questionnaires or tests? If so you will be asked to provide a sample copy of the survey or test in PDF Format)*

Risks: *(What are the risks or potential discomforts: physical, psychological, social, legal, etc.?)*

Minimizing Risk: *(How will you minimize or reduce any potential risk?)*

Benefits: *(How will this study be beneficial?)*

Privacy: *(Will you collect any personal data from your subjects (name, address, photos, etc.?)*

Information Storage: *(How will you store and protect the participant's private information?)*

Information Disposal: *(What will you do with the participant's personal data at the end of the study?)*

VERTEBRATE ANIMAL STUDIES (*complete if applicable, delete if not needed*)

Alternatives: *(Discuss possible alternatives to vertebrate animal use and justify why your project requires the use of vertebrate animals.)*

Kind of Animal: *(Give the common name or genus and species of the animals you will be using)*

Description: *(Describe the animals you will be using, sex, age, etc.)*

Numbers: *(Give the number of vertebrate animals that will be involved in your study and why you chose that number)*

Source: *(Where and how will you obtain the animals?)*

Procedure: *(Describe in detail the exact procedure you will be using on the vertebrate animals)*

Dosages: *(Give detailed chemical concentrations or drug or food dosages that you will be using if applicable)*

Discomfort: *(Explain the ways you will minimize pain and/or discomfort for the animals)*

Husbandry: *(Describe how and where the animals will be kept. Include bedding, cage, type of food, frequency of food and water etc.)*

Disposition After Experimentation: *(What will happen to the animals when you are done with your study?)*

<u>Note: Licensing and other Documentation</u> *(if applicable)***:**

(You will be required to provide a copy of any licensing required by state, local or federal law i.e. fishing license etc. if applicable, in PDF Format)

(You will be required to provide a copy of the IACUC documentation in PDF Format if the study was performed in a research institution. The student name must be on the IACUC or on a letter from the IACUC showing that the student was trained and authorized to work on the project.)

POTENTIALLY HAZARDOUS BIOLOGICAL AGENTS (*complete if applicable, delete if not needed*)

Projects in this section can NOT be done at HOME!

Biosafety Level: *(What have you determined your Biosafety Risk Level to be? BSL-1 or BSL-2)*

Biosafety Determination: *(How did you determine the biosafety level of this project?)*

Identification: *(Identify the potentially hazardous biological agents or micro-organisms to be used. Include the source, catalog number, quantity etc. and biosafety level group for each microorganism)*

Laboratory Environment: *(Describe the laboratory and setup and include the level of biological containment "BSL-Level")*

Minimizing Risk: *(Describe the procedures that will be used to minimize risk: protective equipment, gloves, hoods, etc.)*

Disposal: *(Describe the methods of disposal for all cultures and hazardous biological agents (i.e. autoclave, bleach, etc.)*

Vertebrate Animal/Human - Bodily Fluids/Tissue (complete if applicable, delete if not needed)

Identification: *(What vertebrate animal tissue will be used in this study? Check all that apply)*:

☐**Fresh or Frozen Tissue Sample**

☐**Fresh Organ or Other Body Part**

☐**Blood**

☐**Primary Cell/Tissue Culture**

☐ **Human or Other Primate Established Cell Lines**

Source: *(Where will the tissue be obtained? (include source and catalog number if using established cell lines)*

Tissue from Research Institution: *(Complete if the tissue will be obtained from a vertebrate animal study conducted at a research institution)*

- **Title:** *(Type in the title of the study at the Research Institution)*
- **IACUC:** *(Type in the IACUC approval number)*
- **IACUC Date:** *(Type in the IACUC approval date)*
- **Note:** *(You will be required to upload a copy of the IACUC documentation in PDF Format)*

HAZARDOUS MATERIALS/ACTIVITIES/DEVICES (*complete if applicable, delete if not needed*)

Risk: *(Identify and assess the risks involved in this project)*

Safety: *(Describe the safety precautions and procedure you will use to reduce the risk)*

Details: *(List all hazardous chemicals, activities, or devices be sure to include any chemical concentrations or dosages; also identify any microorganisms exempt from pre-approval that you will be using if applicable).*

Sources of Safety Information: *(List all sources of safety information: product labels, teacher, MSDS, etc.)*

CONTINUATION PROJECTS (*complete if applicable, delete if not needed*)

Previous Year: *(Select the previous year's research)*

☐ **2022-2023 Season**

☐ **2021-2022 Season**

Previous Year's Project Title:

Research Goal - Current Year:

Research Goal – Previous Year:

Method/Procedure Summary – Current Year:

Method/Procedure Summary – Previous Year:

Variables to be Tested – Current Year:

Variables Tested – Previous Year:

Additional Changes/Differences – Current Year:

Additional Changes/Differences – Previous Year:

Note: *(You will be asked to provide a copy of previous year's Research Plan & Abstract in PDF format)*