MATH 750 / DATA 750 / CMPS 750 - DMC Master's Thesis

Project Proposal Template

Important Notes:
- This proposal must be approved by the thesis advisor and two readers prior to registering for MATH/DATA/CMPS 750.
- In order for a proposal to be approved prior to registration, the student must have started a dialog between the advisor and readers at least 4 weeks prior to registration. It is recommended that students initially contact (select) their advisor during the early part of the preceding semester to begin the discussion.
- Projects that require the collection or evaluation of personal information may require approval from the College's Institutional Review Board (IRB) - which must be applied for prior to registration.
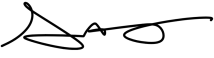
Student Name: Anit Mathew

Semester of Thesis: 4TH

Thesis Advisor: Prof. Osei Kofi Tweneboah

Signature   *Osei Tweneboah*   Date 08/06/2023

Thesis Reader 1:   Prof. Scot Frees

Signature   Date 08/06/2023

Thesis Reader 2:   Prof. Sourav Dutta

Signature   Date 08/06/2023

Proposal Narrative

Please answer each of the following questions in no more than 300 words (each):

1. **What problem are you trying to solve?**

   Explore and examine the connection between aging, lifestyle choices, cardiovascular diseases (CVDs), and brain strokes. With regard to CVDs and brain strokes in the elderly population, I want to specifically look into how lifestyle factors like gender, age, hypertension, heart disease, marital status, type of work, type of residence, average glucose level, body mass index (BMI), smoking status, and stroke are connected. Understanding how these lifestyle factors affect the incidence and prevalence of CVDs and brain strokes in older people is the aim of this research.

2. **Why is the problem worth solving? What industry, research question, or group would benefit from the outcome of the thesis project?**

   Based on existing data, I believe there is an association between lifestyle factors, cardiovascular disease (CVD) and stroke to a certain extent, in older adults which is worth solving for several reasons.

   Public Health Impact:

   Cardiovascular disease and stroke are major health risks, especially for older people. Understanding the relationship between lifestyle factors and these diseases will provide valuable insights for prevention, early detection and targeted interventions to reduce the burden of cardiovascular disease and stroke in this vulnerable group. may provide.

   Elderly care and geriatrics:

   Medical professionals such as geriatricians, cardiologists, and neurologists can use the results to develop effective strategies for risk assessment, disease management, and improving the overall health of older adults.

   Security and Health Promotion:

   The results of this study may help policy efforts related to cardiovascular health and stroke prevention, especially in the setting of an aging population. Health promotion campaigns and interventions can be designed to target specific lifestyle factors that influence the development of cardiovascular disease and stroke in older adults.

   Research Foundation:

   This Research project contributes to the existing body of knowledge on the relationship between lifestyle factors, cardiovascular disease, and stroke in older adults. Findings may

serve as a basis for further research, leading to advances in the field and the potential discovery of new interventions and risk mitigation strategies.

Overall, the project results will provide evidence-based knowledge on how to improve cardiovascular health and reduce stroke incidence in older adults to help health professionals, policymakers, researchers, and others. It may ultimately benefit older people.

3. **Why is the problem hard? What are the key challenges that make the project / solution / deliverable impactful?**

The problem of investigating associations between lifestyle factors, cardiovascular disease (CVD), and stroke in older adults raises several challenges that make the project effective.

Complex interaction:

Lifestyle factors such as age, gender, hypertension, and smoking status are interrelated and can interact in complex ways. Understanding the complex relationships and determining their separate and combined effects on cardiovascular disease and stroke requires careful analysis and consideration of potential confounding variables.

Data collection and accuracy:
Accurate and comprehensive data on lifestyle factors, medical history, and stroke incidence in older adults can be difficult to obtain. There may be discrepancies in data collection methods, inconsistent reporting, or missing values that affect the reliability and generalizability of the results.

Longitudinal properties:

Studying associations between lifestyle factors, cardiovascular disease, and stroke in older adults often requires longitudinal data to assess changes over time. Obtaining long-term data can be time-consuming and resource-intensive, and can pose challenges related to participant retention and follow-up.

Causality and vice versa:

Proving a causal relationship between lifestyle factors and the development of cardiovascular disease and stroke is complex. Lifestyle factors can influence the risk of cardiovascular disease and stroke, but the presence of these diseases can also influence lifestyle choices. To draw meaningful conclusions, it is important to address the issue of causality and consider possible reverse causality.

Destructive factors:

Lifestyle factors do not act alone and are influenced by various confounding factors such as socioeconomic status, educational level, access to medical care and genetic predisposition. Controlling for confounding factors and pinpointing the influence of lifestyle factors alone can be difficult.

4. **Does your project require data? If so, what data are you using? Do you have access or permission to use this data?**

   https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brain-stroke-dataset

   Yes, the above dataset is publicly accessed.

5. **What is the specific deliverable?**

   The most important deliverable is a comprehensive research report documenting the study design, methodology, data analysis, results and conclusions. This research paper provides an in-depth empirical investigation of the relationship between lifestyle factors, cardiovascular disease, and stroke in older adults.

   Objectives would be:

   1. Examining the incidence of heart disease, high blood pressure, and brain strokes in seniors.
   2. Examining how aspects of lifestyle, such as smoking status and BMI, affect the frequency of brain strokes.
   3. Evaluating the link between heart disease, high blood pressure, and brain strokes.
   4. Investigating the possible impact of additional variables on the risk of brain strokes, such as gender, average blood glucose levels, and type of housing.
   5. Suggesting interventions and preventative strategies to lower the incidence of brain strokes in the aged population.

6. **What are the ethical considerations around your project? How might various stakeholders benefit from or be harmed by your work?**

   This thesis project, which explores whether there is relationship and to what extent between lifestyle factors, cardiovascular disease (CVD), and stroke in older adults, raises important ethical considerations. These include obtaining informed consent from participants, ensuring the confidentiality and confidentiality of participant information, promoting data sharing and transparency, striving for fairness and impartiality, and prioritizing participant well-being. These include ranking, protection of research integrity, and inclusion of interest groups.

   Obtaining informed consent is important for participants to fully understand the study's purpose, risks, benefits, and implications. Consent should be voluntary, and participants should be able to withdraw their consent at any time without consequences. Privacy and confidentiality must be maintained throughout the research process through data anonymization, secure storage, and

protection of participants' identities. Researchers must transparently communicate their data collection, storage and sharing practices in accordance with data protection regulations. Equity and equity should guide participant selection, avoid discrimination, and consider health inequalities so that results are applicable to diverse populations.

Stakeholders for this thesis project include older participants, health professionals, policymakers, and public health officials. Older participants may benefit from research by gaining insight into lifestyle factors that influence cardiovascular health and stroke risk. Their rights and welfare should be protected through justice. Health professionals will gain a better understanding of these factors, leading to improved risk assessment, preventive strategies, and patient care. Policymakers and public health officials can use this research to support policies, interventions and campaigns aimed at reducing the burden of cardiovascular disease and stroke in older adults. Overall, addressing these ethical considerations and involving relevant stakeholders ensures that research is conducted responsibly and benefits both participants and the wider community.

7. **What are the core Mathematical, Data Science, and/or Computational skills that you will need to complete your thesis?**

These skills are essential for data analysis, modeling, and drawing meaningful conclusions.

Statistical analysis:

Knowledge of statistical analysis is essential for exploring relationships between variables, identifying trends, and drawing valid conclusions from data. Knowledge of descriptive statistics, hypothesis testing, regression analysis, and survival analysis (where applicable) will be helpful.

Data preprocessing and cleansing:

The ability to preprocess and clean up datasets is critical to ensuring data quality. Knowledge of cleaning data, handling missing values, handling outliers, and standardizing or transforming variables is required.

Exploratory Data Analysis (EDA):
With the help of EDA technology, data insights are gained, patterns are recognized, and relationships are visualized. Knowledge of data visualization, data summarization, and graphing helps you understand the properties of data sets.

Machine learning.

Knowledge of algorithms and machine learning techniques is critical for building predictive models and studying the relationship between lifestyle factors and cardiovascular disease/stroke. Knowledge of classification algorithms (logistic regression, decision trees, random forests, gradient boosting, etc.) and model evaluation is essential.

Programming and data manipulation:
Data manipulation, implementing statistical analysis, and developing machine learning models

require knowledge of programming languages such as Python and R. Familiarity with data manipulation libraries (Pandas, Dplyr, etc.) and machine learning frameworks (Scikit-Learn, TensorFlow, etc.) is an advantage.

Arithmetic Skills:
Large amounts of data and complex analyzes can require computational skills to use computational resources efficiently. Knowledge of parallel computing, cluster computing, and optimizing code for performance helps scale analytics and improve computational efficiency.

Study design and methodology:
Understanding the principles of study design, experimental design, and methods related to the research question is important for designing studies, choosing appropriate statistical methods, and ensuring the validity and reliability of results.

Data ethics and privacy:
When conducting research involving human subjects or sensitive health data, it is important to be familiar with ethical considerations related to privacy, confidentiality, and data management.

8. **What are the specific milestones you expect to accomplish, and when? Your proposal should have a project plan, outlining specific steps you will take to complete the project and a timeline for the work.**

Weeks 1-2:

A Comprehensive review of relevant literature on lifestyle factors, cardiovascular disease, and stroke in older adults. Familiarize with the dataset and narrow down our research objectives and specific research questions.

Weeks 3-4:
Preprocess data by handling missing values, detecting outliers, and cleaning data sets. Conduct descriptive analyses to examine the incidence of heart disease, hypertension, and stroke in older adults.

Weeks 5-6:
Using appropriate statistical methods, investigate the relationship between lifestyle factors (smoking, BMI, etc.) and the incidence of stroke. We will assess the effect of additional variables (gender, mean blood glucose level, residence type) on stroke risk.

Weeks 7-8:
Investigate the relationship between heart disease, hypertension, and stroke using statistical analysis. Identify significant associations or correlations between these variables.

Weeks 9-10:
We use machine learning techniques to analyze datasets and build stroke prediction models based on lifestyle factors and other variables. Evaluate model performance and determine the importance of different features in predicting stroke. Weeks 11-12:
Summarize the results and interpret the results in relation to our research question and goals. Identify patterns, trends, and potential risk factors associated with stroke in older adults.

Weeks 13-14:

To develop recommendations for interventions and prevention strategies to reduce the incidence of stroke in older adults. Write the research paper with all required sections.

Weeks 15-16:

Complete research paper with attention to clarity, consistency and correct references. Prepare visualizations or supporting material to support the findings. We conduct a thorough review and proofread the final document.