VinUniversity Data Science Mini-contest: Innovating with Data Science
Submission Guidelines

I.   Introduction

The second round of the mini-contest is titled **"Innovating with Data Science"**, focuses on developing innovative solutions to real-world challenges prevalent in Vietnamese society, spanning areas such as education, health, tourism, agriculture, manufacturing, and sustainability.

Your task for this round is to identify a pressing socio-economic issue in Vietnam and propose an innovative approach to tackling the problem using data science methodologies. You are to present a proposal outlining the problem, potential data sources and data science methodologies, expected outcomes, and potential impact. This involves not just conceptualizing a data-driven approach to tackle the problem, but also effectively communicating your idea through a written project proposal and a compelling video presentation. The guidelines below provide instructions on how to craft your proposal and present your project idea. Remember, your submission should reflect both your understanding of data science and your vision for its application in addressing real-world challenges in Vietnam. Prototypes are not expected, nor is any preliminary data analysis or data visualisation. This is a proposal about a problem only.

After evaluating the proposal and video, the top 10 most outstanding ones will proceed to the online pitching session where they will pitch their ideas directly to a panel of judges.

II.   Contest Guidelines

II.1. Written Proposal
II.1.a. Format
- **Length**: The proposal should be concise, with a maximum of **2 pages** not including references, which may be on a third page.
- **Formatting:** A PDF document with reasonable font size and margins.

II.1.b. Content
Your proposal **must** include the following:
- **Executive summary:** A succinct overview of the challenge, your data science solution, and the anticipated impact.
- **Problem statement:** Clearly articulate the specific Vietnamese socio-economic issue being tackled and why it is a problem.
- **Proposed approach:**
    - Indicate the type of data (publicly available, collected, etc.), potential sources, and a brief description of possible data processing pipelines and analysis.
    - Detail how data science should be utilized to address the problem.
    - Any necessary developmental steps required for your project.
- **Expected outcomes:** State some measurable results you hope to achieve.
- **Potential impact:** Discuss the prospective societal or industrial benefits.
- **References:** Cite any sources referenced in your proposal.

II.2. Video

II.2.a. Format
- **Length:** The video should not exceed 4 minutes.
- **Technical specifications:**
    - The video should be submitted in MP4 format.
    - The video should have reasonable visual clarity with at least 720p resolution.
    - The audio should have presenter's speech in English, be clear and easy to understand with minimal noise/distractions.

II.2.b. Content
- **Team introduction:** Brief introduction of team members and their roles.
- **Project:** Present the problem and your proposed solution engagingly and persuasively.
    - Use visual aids like slides or infographics to support your explanation.
    - Ensure key points are communicated clearly and concisely.
- **Conclusion:** Briefly summarize the impact and contribution of the project.

II.3. Online Pitching Session
The 10 most outstanding teams will be selected to advance to the prestigious online pitching session. These teams will have the opportunity to pitch their ideas live online and receive direct feedback from the panel of judges.

III.   Evaluation Criteria
Submissions will be evaluated based on a balanced approach that values both public engagement and the expert assessment of the project's quality. This dual approach ensures a fair and comprehensive evaluation of each submission, recognizing both innovative ideas and effective communication skills.

III.1. Written Proposal and Video
III.1.a. Audience Engagement Metrics (40%)
Engagement metrics are crucial for assessing the public's interest and interaction with the project proposals. The following points will be awarded based on audience engagement with the video presentations:
- **Like:** 1 point for each like.
- **View:** 3 points for each view.
- **Share**: 5 points for each share.

III.1.b. Assessment by Organizers/College of Engineering and Computer Science (60%)
The remaining 60% of the score will be determined by an expert panel of judges, including members from the organizing body and the College of Engineering and Computer Science.

III.2. Online Pitching Session
This assessment will focus on three key areas:
- Project Idea Presentation:
    - Effectiveness in presenting the data science project, including a clear explanation of the problem, solution, expected outcomes, and potential impact.
    - Ability to convey the project's significance within the allocated time frame.
- Innovation and Practicality:

- - Originality of the idea: How the project stands out in terms of creativity and uniqueness.
  - Practicality: Feasibility of the proposed solution and its applicability in addressing the identified issue.
  - Societal or industry impact: Potential benefits and real-world applicability of the project.
- Presentation Skills:
  - Delivery: The clarity and coherence of the presentation.
  - Engagement: Ability to maintain audience interest and interact effectively.
  - Complexity Communication: Skill in explaining complex data science concepts in an accessible manner.

IV.   Example Project Ideas.

The following examples give brief examples of the sorts of content we would like to see. Feel free to build on one of these as you own submission if you like, since a lot more detail is required.

- **Air Quality and Environmental Pollution**: Vietnam faces significant challenges with air pollution, especially in major cities like Hanoi and Ho Chi Minh City, affecting health and quality of life.
  - Data: Air quality indices, pollution levels (PM2.5, PM10, NO2, etc.), health data, traffic data.
  - Data science solutions: Analyzing pollution sources, predicting air quality levels, impact assessment on public health, and policy evaluation.
  - Potential impact: Improved air quality leads to better public health outcomes, reduced healthcare costs, increased productivity, and enhanced quality of life for residents. Additionally, effective policies informed by data science solutions can drive sustainable development and attract investment in cleaner industries.

- **Agricultural Productivity and Climate Change**: Agriculture is a key sector in Vietnam's economy, but it's vulnerable to climate change impacts like rising temperatures and unpredictable weather patterns.
  - Data: Weather data, crop yield data, satellite imagery, soil health data.
  - Data science solutions: Predictive models for crop yields, climate impact studies, optimization of water and resource usage, identifying areas at risk of climate change effects.
  - Potential impact: Enhanced agricultural productivity mitigates food insecurity, boosts farmer incomes, and strengthens rural economies. By adapting to climate change through data-driven strategies, farmers can minimize losses, maintain stable livelihoods, and contribute to national food security and economic resilience.

- **Education and Learning Outcomes in Rural Areas**: Enhancing educational outcomes and addressing disparities in access to quality education across different regions.
  - Data: Student performance data, demographic data, online learning engagement data.
  - Data science solutions: Analysis of factors influencing learning outcomes, predictive models for at-risk students, resource allocation optimization.
  - Potential impact: Improving educational outcomes in rural areas fosters social mobility, reduces poverty, and narrows socioeconomic disparities. Enhanced access to quality education empowers individuals and communities, driving economic development, innovation, and social cohesion. Additionally, optimizing resource allocation ensures efficient use of public funds and maximizes the return on educational investments.

- **Empowering rural communities with digital literacy and technology access**: Rural areas in Vietnam lack sufficient digital literacy resources and technology access compared to urban regions. This gap limits rural residents' participation in the digital economy and access to essential services, worsening socioeconomic disparities. Bridging this divide is crucial for fostering rural development, economic empowerment, and social inclusion in an increasingly digital world.
    - Data: Survey data on digital literacy levels, internet access data, mobile phone usage data.
    - Data science solutions: Clustering analysis to identify vulnerable communities; natural language processing to translate educational materials into local languages; machine learning to predict demand for specific digital skills.
    - Potential impact: Improved access to information and services, increased employability, economic development in rural areas.