**CAPSTONE PROJECT GUIDELINE FOR EACH SECTION**

**SECTION 1: INTRODUCTION**

1. Multi-Source Data. Eg weather, soil & Plant growth data.
2. Machine Learning Algorithms including supervised, unsupervised & reinforcement learning to enhance prediction of changes.
3. Considering technologies such as hyperspectral imaging and remote sensing.
4. Precision Agriculture: (Analytics & IoT) GIS, Remote sensing.
5. Confidence calibration in neural networks & investigate practical solutions for improving confidence estimates.
6. Research of semantic meaning of individual units in deep neural networks & existence of adversarial examples.
7. Evaluate & compare NoSQL database to determine the most suitable option for storing & managing agricultural data.

**SECTION 2: OBJECTIVES**

Problem Definition: Development of a scalable & accurate system for automated crop disease diagnosis that integrate machine learning & image identification techniques.

1. Deep Learning model using Multi-Source Data
2. Investigate the effectiveness of different NoSQL databases for storing Agricultural data.
3. Explore different deep learning paradigms to enhance the performance of models.

**SECTION 3: VALIDITY MANAGEMENT**

1. Accurate
2. Relevance
3. Reliable

Image, Weather data, soil data

**SECTION 4: SAMPLING STRATEGY**

Non-probability sampling strategy. (Snowballing & Judgement)

1. It is essential to gather information from experts and researchers in the field.
2. Snowballing:- Can get these experts through referrals or recommendations.
3. Judgement sampling:- Using judgement to select participants based on their relevance, expertise and ability to contribute to the project.

**SECTION 5: PRIMARY RESEARCH METHODOLOGY**

1. IN-DEPTH INTERVIEWS (Qualitative Research Method)

Questions for the In-depth Interviews

In-depth Interview Questions for Farmer/Agricultural Scientists:

1. Can you describe your experience with crop diseases in your agricultural practices?
2. How do you currently diagnose and manage crop diseases on your farm?
3. What are the major challenges you face when identifying and treating crop diseases?
4. How do you make decisions regarding the use of pesticides or other treatments for crop diseases?
5. What weather conditions or environmental factors have you observed to be correlated with the occurrence of crop diseases?
6. Have you tried using any automated or technological solutions for crop disease diagnosis?
7. What types of data (e.g., weather data, soil data, images) do you think would be most useful for diagnosing and managing crop diseases effectively?
8. How would an automated crop disease diagnosis system benefit your farming practices and what are your expectations from such a system?
9. How would you envision integrating an automated crop disease diagnosis system into your daily farming routine? Are there any challenges or concerns you would expect?
10. In your opinion, what would be the most important features or functionalities of an automated system for crop disease diagnosis, considering the practicality and usability on the farm?
11. How do you currently access and manage agricultural data, such as weather information or soil data? Are there any difficulties or limitations you face in this regard?
12. What level of expertise or technical knowledge do you think would be required for farmers to effectively use an automated crop disease diagnosis system? Are there any concerns about the learning curve or usability?
13. Are there any specific concerns or considerations regarding data privacy or data security that you would have with an automated crop disease diagnosis system?
14. what would be the potential impact of an accurate and efficient automated crop disease diagnosis system on your farm's productivity, costs, and sustainability practices?

In-depth Interview Questions for Data Scientists:

1. Could you describe your experience working with multi-source data integration for crop disease diagnosis or similar applications?
2. What are the key challenges or limitations you have encountered when utilizing deep learning techniques for crop disease diagnosis? How have you addressed or mitigated these challenges?
3. In your opinion, what are the most suitable deep learning algorithms or architectures for accurate and efficient crop disease diagnosis based on image identification? Can you explain why?
4. How have you utilized machine learning algorithms, such as supervised, unsupervised, or reinforcement learning, to enhance the prediction of changes and optimize farming practices in the context of crop disease diagnosis?
5. Can you discuss any specific research or advancements in deep learning paradigms that have shown promise in improving crop disease diagnosis accuracy and efficiency?
6. What are the considerations or trade-offs when integrating different data sources, such as weather data, soil data, and images, for crop disease diagnosis? How do you handle data preprocessing and fusion?
7. Have you encountered any challenges in calibrating confidence estimates in deep neural networks for crop disease diagnosis? How have you approached this issue, and what practical solutions have you found?
8. Can you share any insights or findings regarding the semantic meaning of individual units in deep neural networks and their implications for crop disease diagnosis? How does this contribute to model interpretability and reliability?
9. Have you explored the existence of adversarial examples in crop disease diagnosis models? How do you ensure the robustness and resilience of the models against such examples?
10. What are the critical considerations when selecting and evaluating NoSQL databases for storing and managing agricultural data? What factors do you prioritize in terms of query response time and optimization for specific operations?
11. Based on your experience, how do you envision the integration of machine learning and IoT technology in precision agriculture for crop disease diagnosis? Are there any potential challenges or limitations to be addressed?
12. Can you provide examples of successful implementations or applications of machine learning algorithms, such as multilayer perceptron rules-based classifiers or decision table classifiers, for crop disease diagnosis in real-world scenarios?
13. How would you suggest integrating the outputs of an automated crop disease diagnosis system into the decision-making process of farmers? What form or format of recommendations would be most effective and actionable?
14. Are there any considerations or challenges related to data privacy, security, or ethics when developing and implementing an automated crop disease diagnosis system?
15. Based on your expertise, what impact do you anticipate an accurate and efficient automated crop disease diagnosis system could have on farming practices, crop yield, and the agricultural industry as a whole?

**SECTION 6: ETHICAL CONSIDERATIONS**

1. Robust Data Encryption
2. Anonymizing Personally Identifiable Information
3. Establishing secure storage & transmission protocals
4. Potential Bias in the Deep Learning models.
5. Transparency and Interpretability

**SECTION 7:** **RESEARCH DESIGN AND METHODOLOGY**

1. Data retrieval and justification for data used.
2. License for data used.
3. Data description and reference explained and justified.
4. Ethical considerations of primary research. Justification of relevance and data license agreement and usage capacity
5. Discussion of the nature of data acquired & justification. Ensure to discuss each section in depth
6. .
7. .
8. .
9. .
10. .
11. .
12. .
13. .
14. .
15. .

**SECTION 8: SEQUENTIAL MODEL COMPILATION PER STAGE**

**Stage 1: Backpropagation Neural Network (BPNN)**

1. **Build Simple BPNN:**
   * A basic BPNN architecture with one hidden layer.
2. **Hyperparameter Exploration:**
   * Explore different learning rates, regularization strengths, and activation functions (e.g., sigmoid, ReLU, tanh).
   * Implement parameter grid search to find optimal hyperparameters.
3. **Validation Techniques:**
   * Apply k-fold cross-validation (e.g., 5-fold) to assess model performance on different data splits.
4. **Visualization and Analysis:**
   * Visualize learning curves showing training and validation loss/accuracy over epochs.
   * Plot confusion matrices to understand the model's classification performance.
   * Compare the effects of different regularization techniques (e.g., dropout, L2 regularization) on performance.

**Stage 2: Artificial Neural Network (ANN)**

1. **Transition to ANN:**
   * Modify the BPNN architecture to include multiple hidden layers.
2. **Hyperparameter Tuning and Exploration:**
   * Experiment with learning rates, dropout rates, number of hidden neurons, and activation functions.
   * Implement parameter grid search to optimize hyperparameters.
3. **Validation Techniques:**
   * Continue using k-fold cross-validation to evaluate model performance.
4. **Visualization and Analysis:**
   * Visualize learning curves and compare them with the BPNN stage.
   * Plot confusion matrices and observe the impact of different activation functions on classification accuracy.

**Stage 3: Convolutional Neural Network (CNN)**

1. **Start with Baseline CNN:**
   * Build a simple CNN architecture with convolutional and pooling layers.
2. **Hyperparameter Tuning and Exploration:**
   * Experiment with learning rates, batch sizes, filter sizes, and pooling sizes.
   * Perform parameter grid search to fine-tune hyperparameters.
3. **Data Augmentation:**
   * Implement data augmentation techniques (e.g., rotation, zoom) to enhance model robustness.(If Necessary)
4. **Transfer Learning:**
   * Explore CNN architectures (e.g., VGG, ResNet, Inception) on a larger dataset.
5. **Validation Techniques:**
   * Continue using k-fold cross-validation to validate CNN performance.
6. **Visualization and Analysis:**
   * Plot learning curves for each CNN variant, including baseline, augmented, and transfer learning models.
   * Visualize confusion matrices to understand class-specific performance.

**Conclusion and Presentation:**

* Summarize the results and insights gained from each stage.
* Discuss the advantages and limitations of each model and technique explored.
* Present a comparison of the performance metrics across stages and models.
* Highlight the final CNN model's accuracy, interpretability, and impact on project.