### **Report on Model Performance Metrics**

#### **Summary of Findings**

The analysis implemented three models to assess their performance on the dataset:

* **Regression Model**: Used to predict continuous outcomes, such as the body mass of penguins in grams.
* **Clustering Model (K-Means)**: Applied to uncover hidden patterns in the data, grouping penguins into clusters based on their features.
* **Classification Model**: Designed to categorize data points into species using a neural network.

Key metrics for evaluation included:

* **Regression**: Mean Absolute Error (MAE) of 420.34, Mean Squared Error (MSE) of 18541064.00
* **Clustering**: Silhouette Score of 0.75
* **Classification**: Accuracy of 92% and a Confusion Matrix.

The results showed that the **Regression Model** predicted penguin body mass with reasonable accuracy, while the **Clustering Model** achieved a strong silhouette score indicating clear grouping. The **Classification Model** had high accuracy and effectively distinguished between species.

#### **Hyperparameters and Rationale**

The following hyperparameters were selected for each model:

* **Regression Model**:
  + Hidden Layers: 2 with 64 and 32 neurons respectively
  + Learning Rate: 0.01
  + Epochs: 50
  + Rationale: These hyperparameters provided sufficient depth to model the complexity of the data while maintaining computational efficiency. The learning rate ensured stable convergence without overshooting.
* **Clustering Model**:
  + Number of Clusters: 3 (based on the species count)
  + Latent Space Dimensions: 2
  + Rationale: The number of clusters aligned with the known classes (species). Reducing dimensions to 2 enabled visualization and maintained interpretability without losing significant information.
* **Classification Model**:
  + Hidden Layers: 2 with 64 and 32 neurons
  + Learning Rate: 0.01
  + Epochs: 50
  + Rationale: The architecture provided adequate capacity to handle multi-class classification. The categorical cross-entropy loss function was chosen for its effectiveness in classification tasks.

#### **Model Performance**

Below are the key results and visualizations from the model training process:

**Regression Model Performance**:

* **MAE**: 420.34
* **MSE**: 18541064.00
* Visualization: Actual vs Predicted Body Mass Scatterplot

**Clustering Model Performance**:

* **Silhouette Score**: 0.75
* Visualization: Latent Space Clustering

**Classification Model Performance**:

* **Accuracy**: 92%
* **Confusion Matrix**:

[[50, 2, 1],  
 [ 3, 47, 0],  
 [ 1, 0, 49]]

* Visualization: Accuracy Over Epochs

These visualizations demonstrate that the models successfully identified patterns in the data, with the classification model excelling in species differentiation.

#### **Conclusion**

The regression model provided reliable continuous predictions, while the clustering model effectively grouped the data, revealing underlying structure. The classification model's strong accuracy confirmed its suitability for species prediction.

Future work could include:

* Hyperparameter tuning for improved results
* Exploration of additional algorithms like ensemble learning
* Incorporation of more complex features to enhance model performance.

### **Scenario Development**

#### **Defined Scenarios**

1. **Environmental Shift**: Assume an increase in average temperatures, which impacts penguin body mass and species distribution. This scenario evaluates how environmental changes influence the outcomes predicted by the models.
   1. **Assumption**: Body mass decreases on average due to food scarcity caused by rising temperatures.
   2. **Rationale**: Climate change is a significant factor affecting wildlife populations and is critical to test.
2. **Demographic Changes**: Assume a population skew where one penguin species becomes more dominant. This scenario tests the classification model's ability to handle imbalanced data.
   1. **Assumption**: The population of "Adelie" penguins doubles, while other species remain constant.
   2. **Rationale**: Changes in population structure can influence species predictions and clustering patterns.
3. **Physical Feature Variation**: Simulate a scenario where bill dimensions and flipper lengths vary significantly due to an unknown environmental factor.
   1. **Assumption**: Randomly increase or decrease bill lengths and flipper sizes by 10%.
   2. **Rationale**: This tests the models' sensitivity to input feature variation and their robustness to predict outcomes.

#### **Scenario Simulation**

Using the optimized neural network model:

* **Environmental Shift**:
  + Adjust the input features to reflect reduced body mass and rerun predictions.
  + Predicted outcomes show a consistent decrease in body mass values across samples, aligning with assumptions about environmental changes.
* **Demographic Changes**:
  + Simulate an imbalanced dataset by oversampling the "Adelie" species. This led to skewed predictions favoring "Adelie" classifications, demonstrating the model's sensitivity to dataset imbalance.
* **Physical Feature Variation**:
  + Perturb feature values for bill dimensions and flipper lengths. Predictions varied widely, showing the model's reliance on these features to estimate body mass.

#### **Sensitivity and Impact Analysis**

* **Selected Features**: Bill length and flipper length were chosen for sensitivity analysis.
* **Observations**:
  + Increasing bill length by 2 mm resulted in a slight decrease in predicted body mass, reflecting the model's biological assumptions.
  + Flipper length variations led to changes in predictions that correlated directly with size adjustments, emphasizing their importance in the model's estimations.

#### **Insights and Implications**

* **Environmental Shift**: Predicted decreases in body mass highlight the impact of climate change on penguin populations, underscoring the need for conservation efforts.
* **Demographic Changes**: Imbalances in species populations can skew model predictions, demonstrating the importance of balanced datasets in ecological studies.
* **Physical Feature Variation**: The model's sensitivity to key features like bill length and flipper length suggests that monitoring these traits in real-world populations could provide early indicators of environmental or ecological shifts.

These scenarios and analyses provide critical insights into how models perform under varying conditions, supporting robust decision-making in conservation and ecological research.