Obesity Prediction Presentation Presenter Script

Slide 1: Introduction to Obesity Prediction

"Good day everyone! Today, I'll be walking you through our Obesity Prediction project. Obesity is a major health concern worldwide, and early prediction can greatly help in prevention and intervention. In this presentation, I'll guide you through the data we used, the models we applied, and most importantly, how these models performed. Let's begin!"

Slide 2: Features - Demographics & Physical Activity

"To predict obesity accurately, we considered two main factors: **Demographics** and **Physical Activity**. Demographics include details like age, gender, and body measurements. Meanwhile, physical activity includes habits like exercise frequency, screen time, and daily movement.

Slide 3: Data Separation of X and Y

"Now, let's talk about data separation. In machine learning, we need to split our dataset into **X** (our input features) and **Y** (our target output). For example, demographics and physical activity go under X, while the obesity status — whether a person is obese or not — goes under Y. This separation is crucial because our model learns patterns from the X data to predict the Y outcomes."

Slide 4: Logistic Regression

"Our first model is **Logistic Regression**. This is a straightforward yet powerful algorithm often used for classification tasks. Logistic Regression works by finding the best curve that separates different classes — in our case, obese versus non-obese individuals. Its simplicity makes it a great starting point for prediction tasks."

Slide 5: Logistic Regression Evaluation

"Now, let's look at how well Logistic Regression performed. We evaluated it using several key metrics:

- **Accuracy** measures the overall correctness.
- **Precision** tells us how well the model avoids false positives.
- **Recall** shows how effectively it identifies true cases.
- And finally, the **F1 Score** balances both precision and recall.

In our results, Logistic Regression showed decent accuracy, but we noticed some trade-offs in precision and recall. This is important to consider when predicting real-world health outcomes."

Slide 6: Decision Tree

"Next, we applied the **Decision Tree** model. This model works like a flowchart, splitting the data based on key conditions. For example, if a person's BMI is high, the model may classify them as obese. Decision Trees are great because they’re easy to interpret, but they can sometimes overfit the data if not carefully tuned."

Slide 7: Decision Tree Evaluation

"Evaluating our Decision Tree model revealed strong performance in some areas, especially for identifying clear patterns in the data. However, we also observed that this model was more prone to misclassifying borderline cases. This highlights the importance of fine-tuning when using Decision Trees."

Slide 8: Random Forest

"Our third model was **Random Forest**. Think of this as a team of Decision Trees working together. Instead of relying on just one tree's decision, Random Forest combines multiple trees and takes a vote. This teamwork reduces the chances of overfitting and improves overall performance."

Slide 9: Random Forest Evaluation

"Random Forest delivered impressive results, showing higher accuracy and better handling of complex data patterns. This model particularly stood out in cases where data points were harder to classify."

Slide 10: Comparison Model - Confusion Matrix (NAAY EXPLANATION SA PINAKA LAST) sa mismong picture nga explanation

"Now, let's compare our models using the **Confusion Matrix**. This is a powerful tool that shows how well each model distinguishes between true and false predictions.

Here's a quick explanation of the matrix:

- **True Positive (TP):** Correctly predicted obese individuals.
- **True Negative (TN):** Correctly predicted non-obese individuals.
- **False Positive (FP):** Incorrectly predicted as obese (when they're not).
- **False Negative (FN):** Missed an actual obese case.

Understanding these values helps us see which model is both accurate and reliable."

Slide 11: Analysis of Confusion Matrix Results

"Now for the results! Comparing the confusion matrices:

- The **Random Forest** model had the most **True Positives**, showing its strength in correctly identifying obese individuals. It also showed the lowest number of **False Negatives**, meaning it rarely missed an actual obese case.
- The **Logistic Regression** model performed well but had a slightly higher number of **False Negatives**, meaning it missed some cases — a concern for real-world scenarios where early detection is crucial.
- The **Decision Tree** model was decent but had more **False Positives**, meaning it sometimes wrongly classified healthy individuals as obese. This could lead to unnecessary concern or intervention.

Overall, **Random Forest** proved to be the most balanced and effective for this dataset."

Slide 12: Conclusion

"To wrap up, Random Forest emerged as the top performer in our obesity prediction task. However, each model has its strengths and weaknesses. Logistic Regression offers simplicity, Decision Trees are easy to interpret, and Random Forest excels in accuracy and robustness. Choosing the right model depends on your goals — whether you prioritize interpretability or predictive power.

Thank you for your attention, and I'm happy to answer any questions!"

For confusion matrix(for further discussion) optional(for understanding lang)

In the confusion matrix image you provided, here's how you can identify the key elements:

Logistic Regression Example:

- **True Positive (TP):** Values on the diagonal from top-left to bottom-right (e.g., 103 for 'Normal', 100 for 'Overweight', etc.). These are correct predictions for each class.
- **True Negative (TN):** This requires adding all the correctly predicted values *outside* the target class's row and column. For example, for 'Obesity', TN would include all correctly classified values for the other three classes.
- **False Positive (FP):** Values in the column of the predicted class but outside the diagonal. For example, for 'Obesity', FP would include entries where 'Obesity' was predicted but the true class was something else.
- **False Negative (FN):** Values in the row of the true class but outside the diagonal. For example, for 'Obesity', FN would include cases where 'Obesity' was the actual class, but the model predicted something else.

Key Observations:

- In the **Logistic Regression** matrix, the model performed well for identifying 'Normal' and 'Overweight' but struggled with correctly classifying 'Obesity'.

- In the **Decision Tree** matrix, the model shows more False Positives, particularly for 'Obesity' and 'Overweight'.

- In the **Random Forest** matrix, there's a noticeable improvement in reducing False Negatives and maintaining strong True Positive values.

By analyzing these details, you can clearly see that **Random Forest** is the most balanced model, effectively reducing both False Positives and False Negatives.