

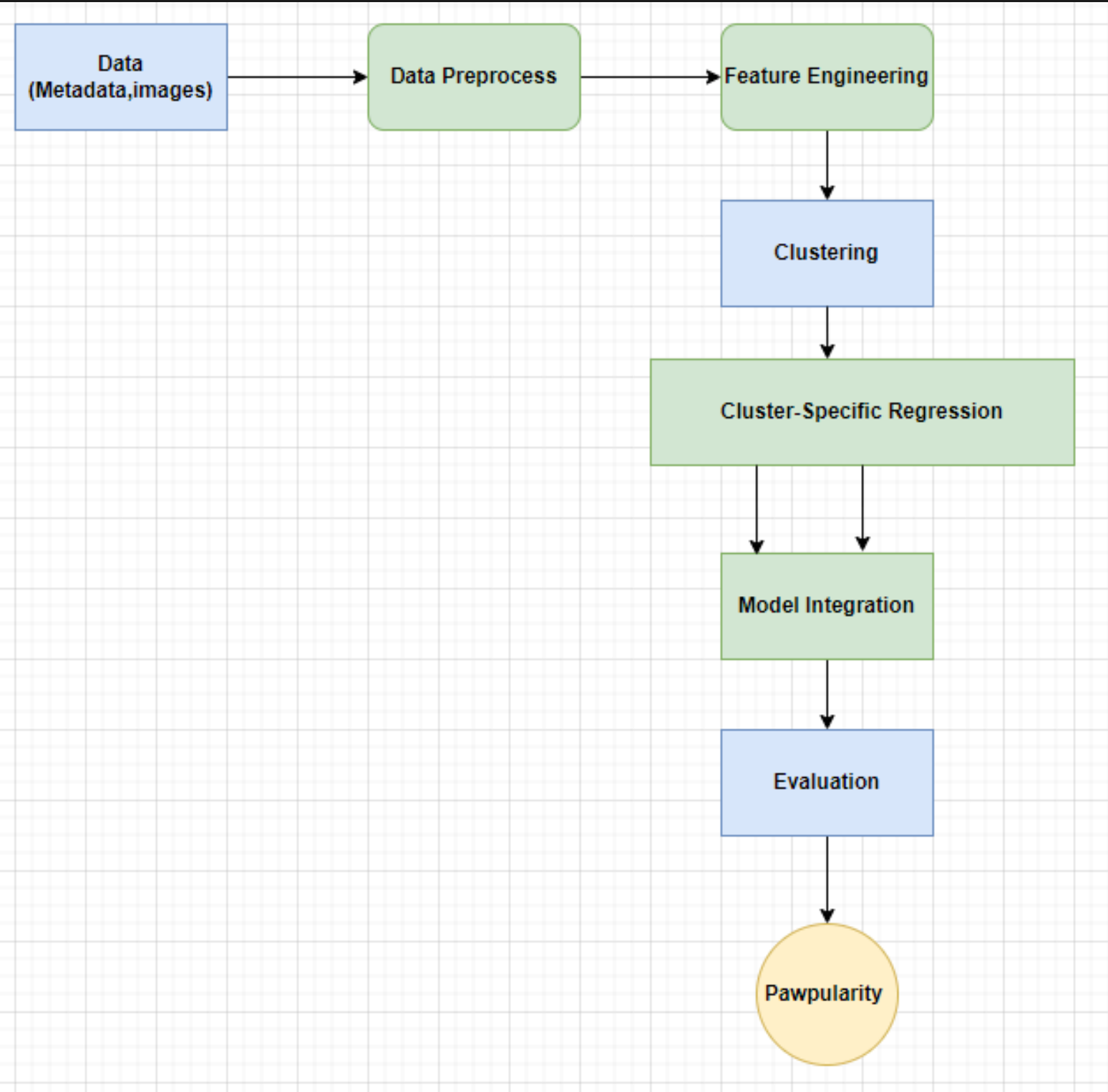
The background is a dark grey or black canvas filled with various colorful, hand-drawn style lines and shapes. These include straight lines, curves, loops, and geometric forms in shades of blue, green, orange, yellow, pink, and purple. Some shapes resemble stylized icons, such as a document with lines, a speech bubble, and a person silhouette. The overall effect is a vibrant, abstract pattern.

Pawpularity Final Project Proposal

GROUP 17

109006271 Tuguldur Ts.

FLOW CHART



01

Data Collection: Gather all relevant data, including images and metadata.

02

Data Preprocessing: Clean and normalize data to ensure it is suitable for analysis.

03

Feature Engineering: Extract and combine features from both images and metadata.

04

Clustering: Apply clustering algorithms to segment the data into similar groups.

05

Cluster-Specific Regression: Develop and train a unique regression model for each cluster.

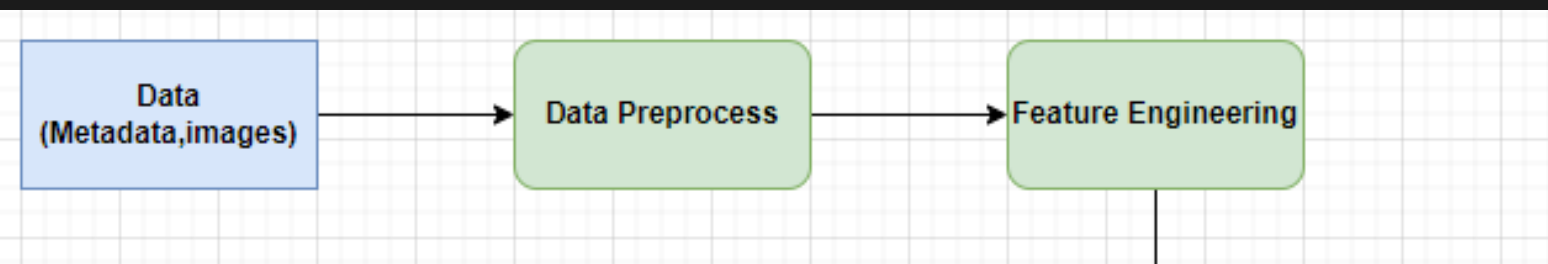
06

Model Integration and Ensemble: Combine the predictions from each cluster-specific model using ensemble techniques.

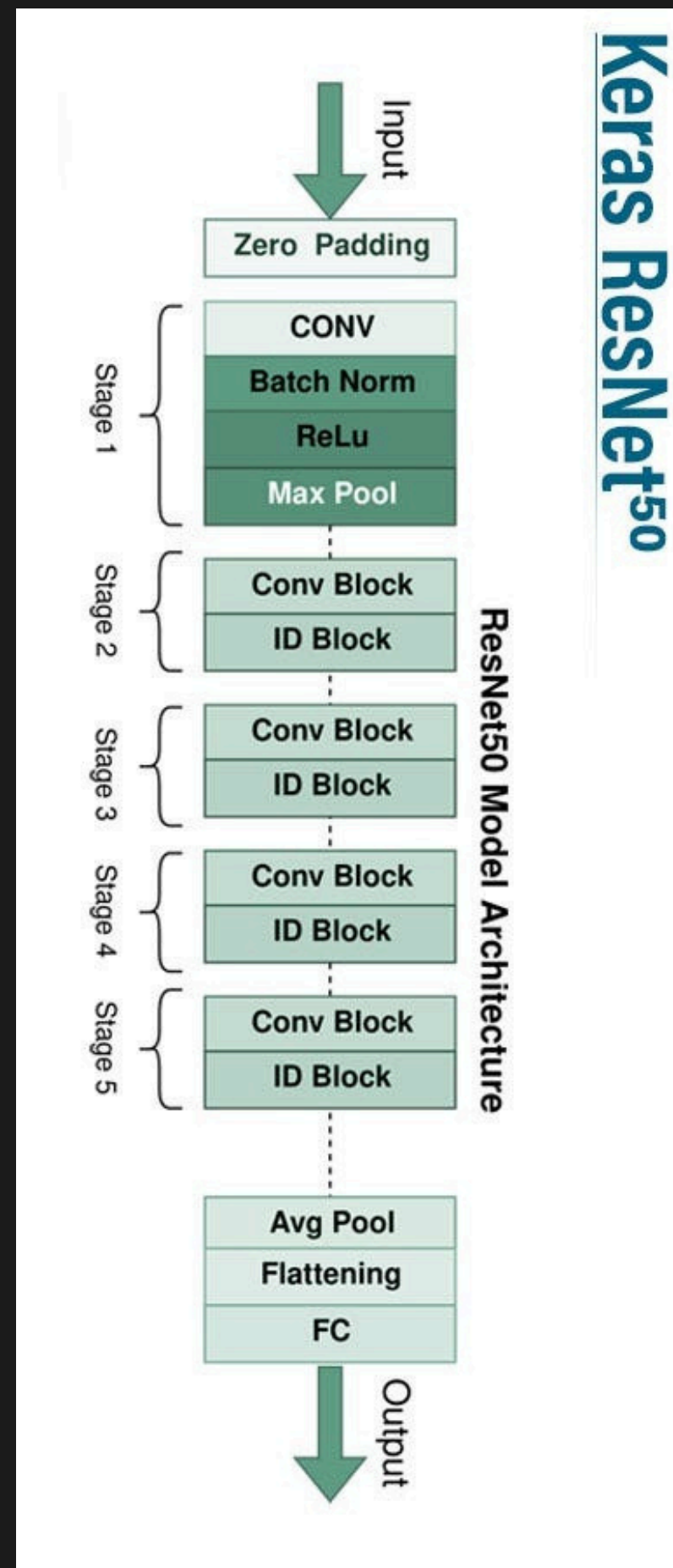
07

Evaluate and Deploy: Deploy the final model for predicting Pawpularity scores on new data.

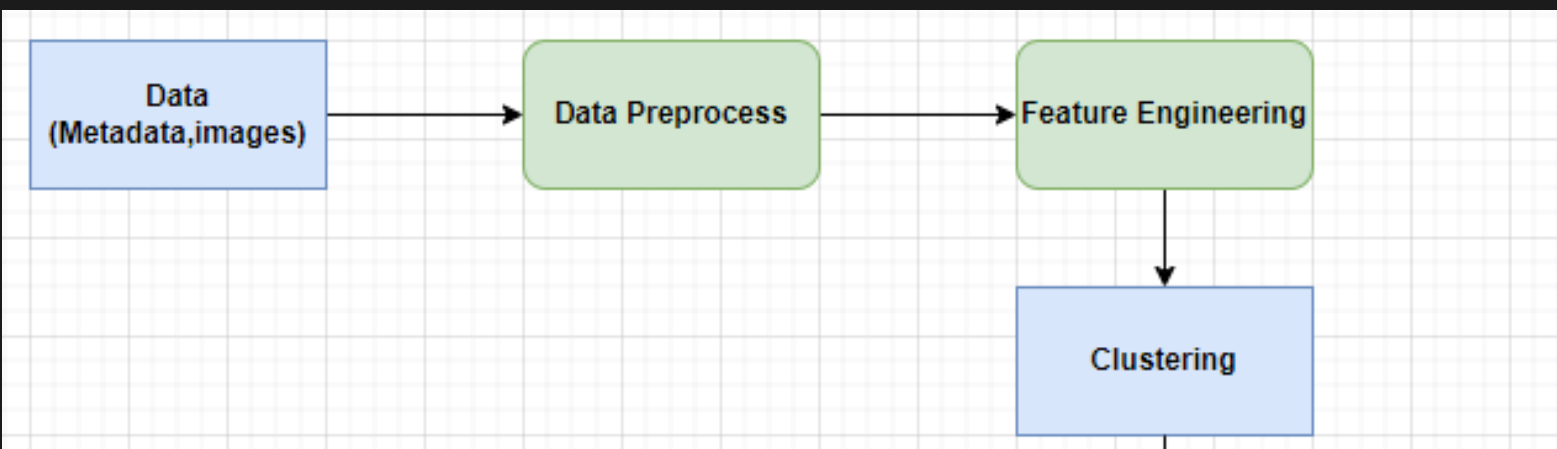
03 Feature Engineering



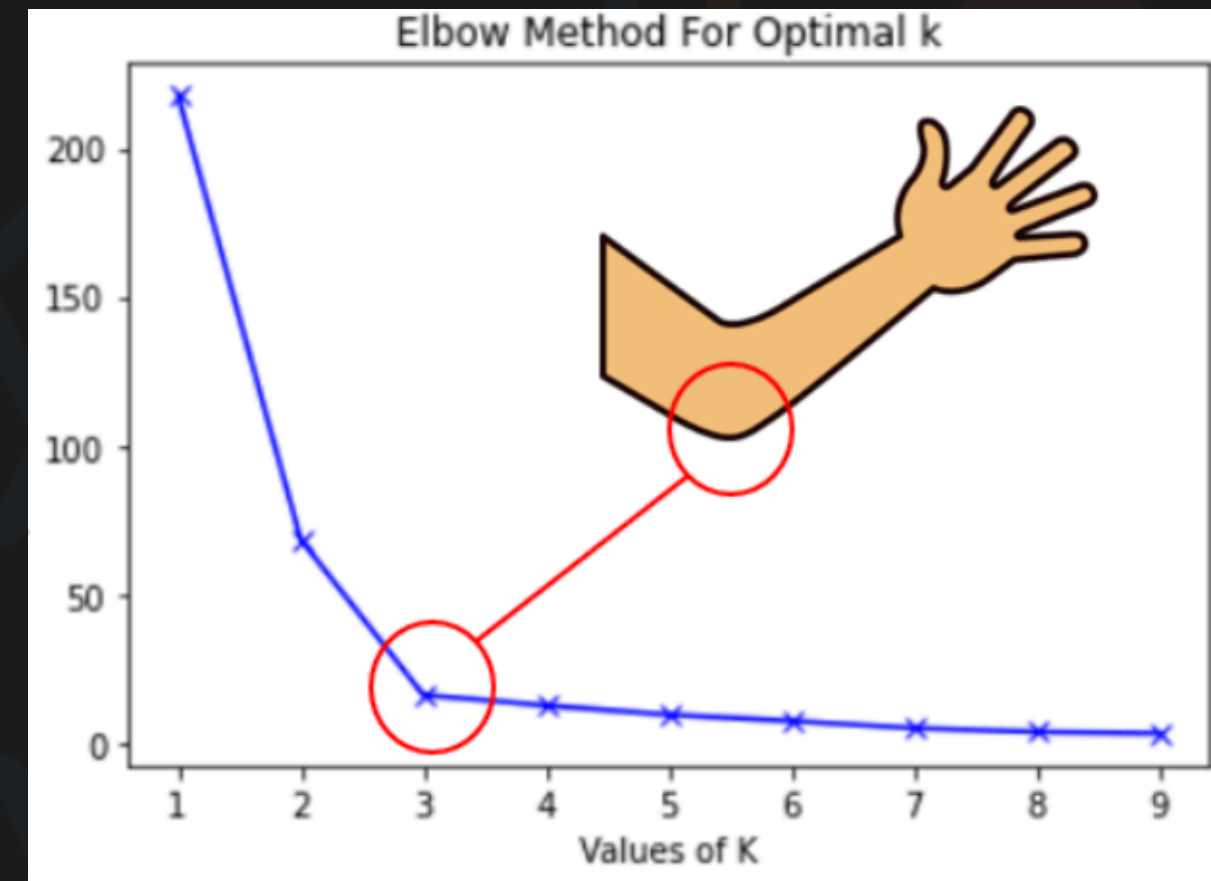
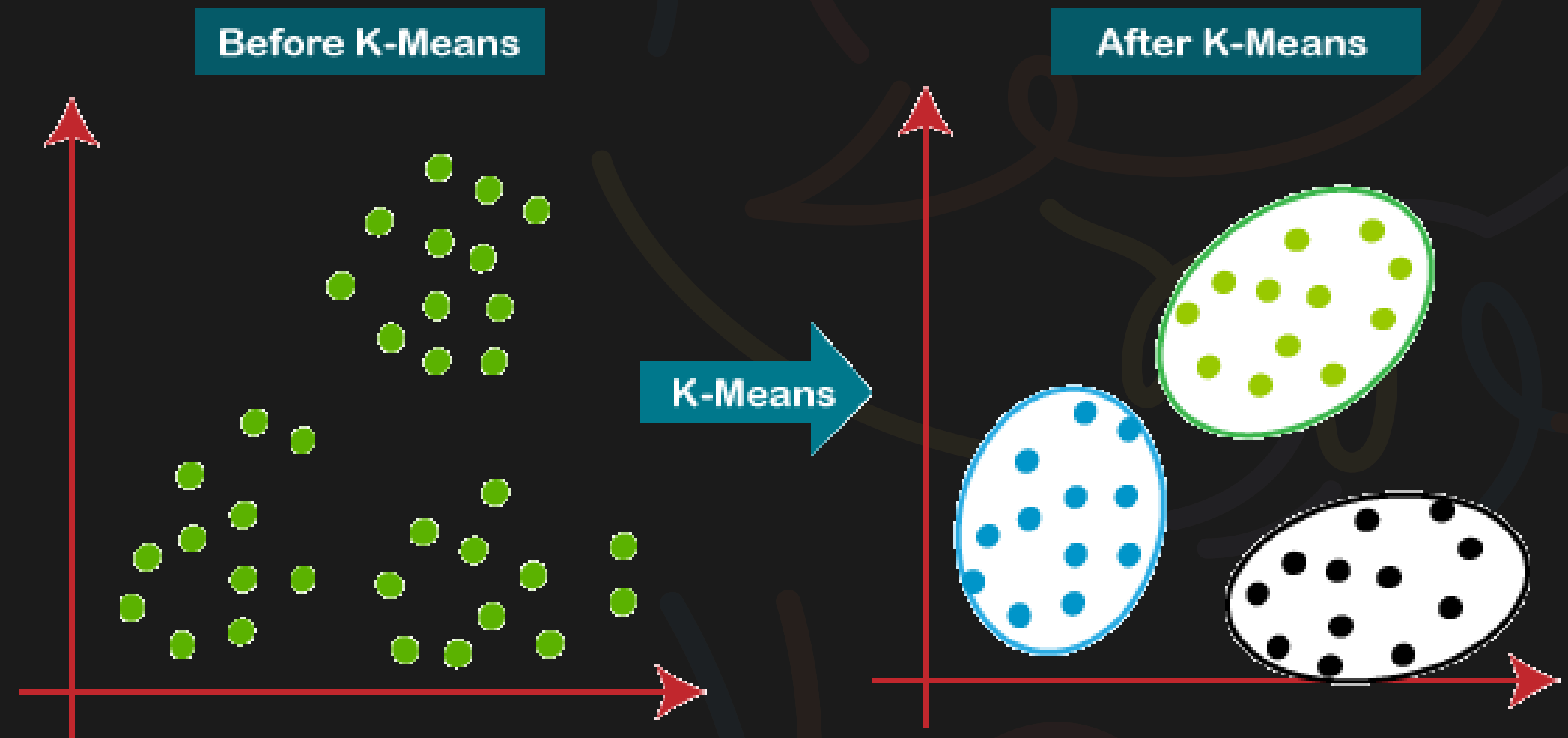
- In feature engineering, I want to employ pre-trained neural networks like ResNet-50 to extract high-level visual features from the images.
- These features are then combined with the metadata to create a feature set that captures both visual and contextual information.
- If I have more extra time, I will try to use another pre-trained model like VGG-16.



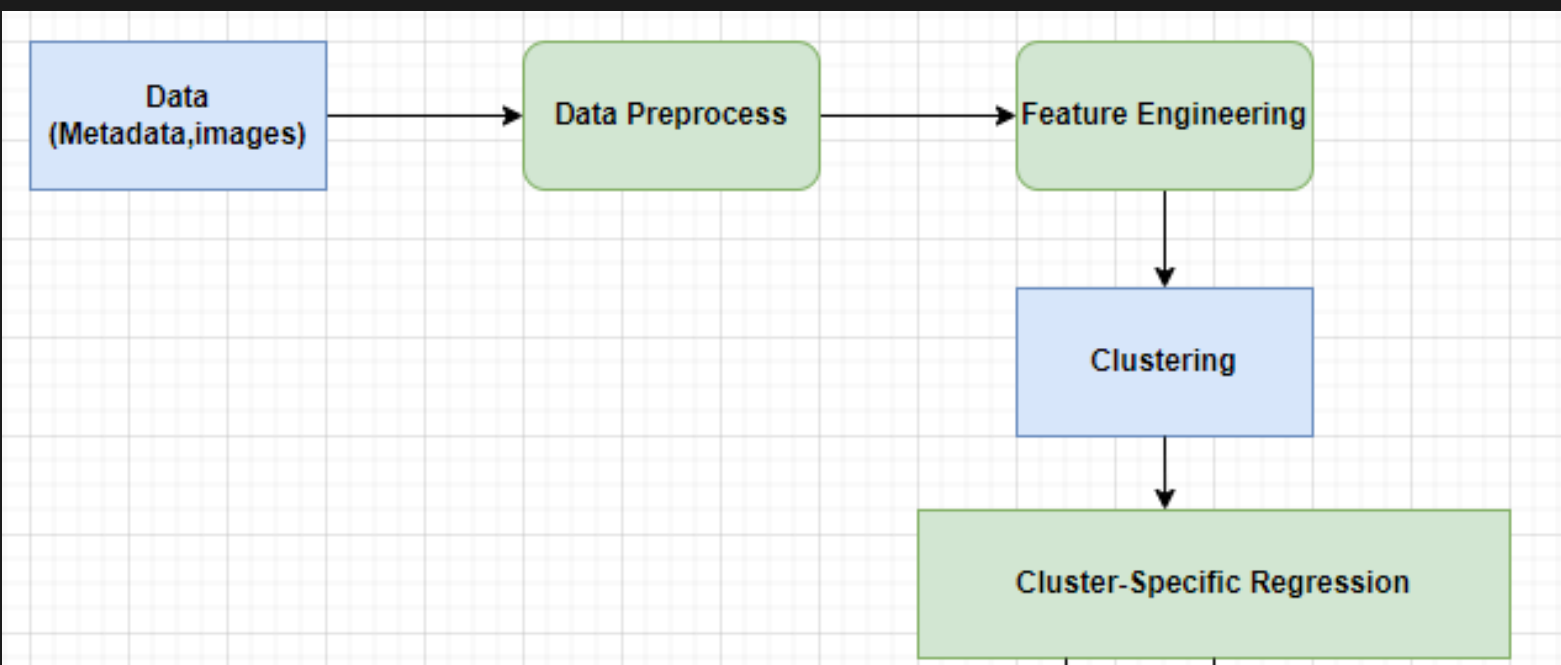
04 Clustering



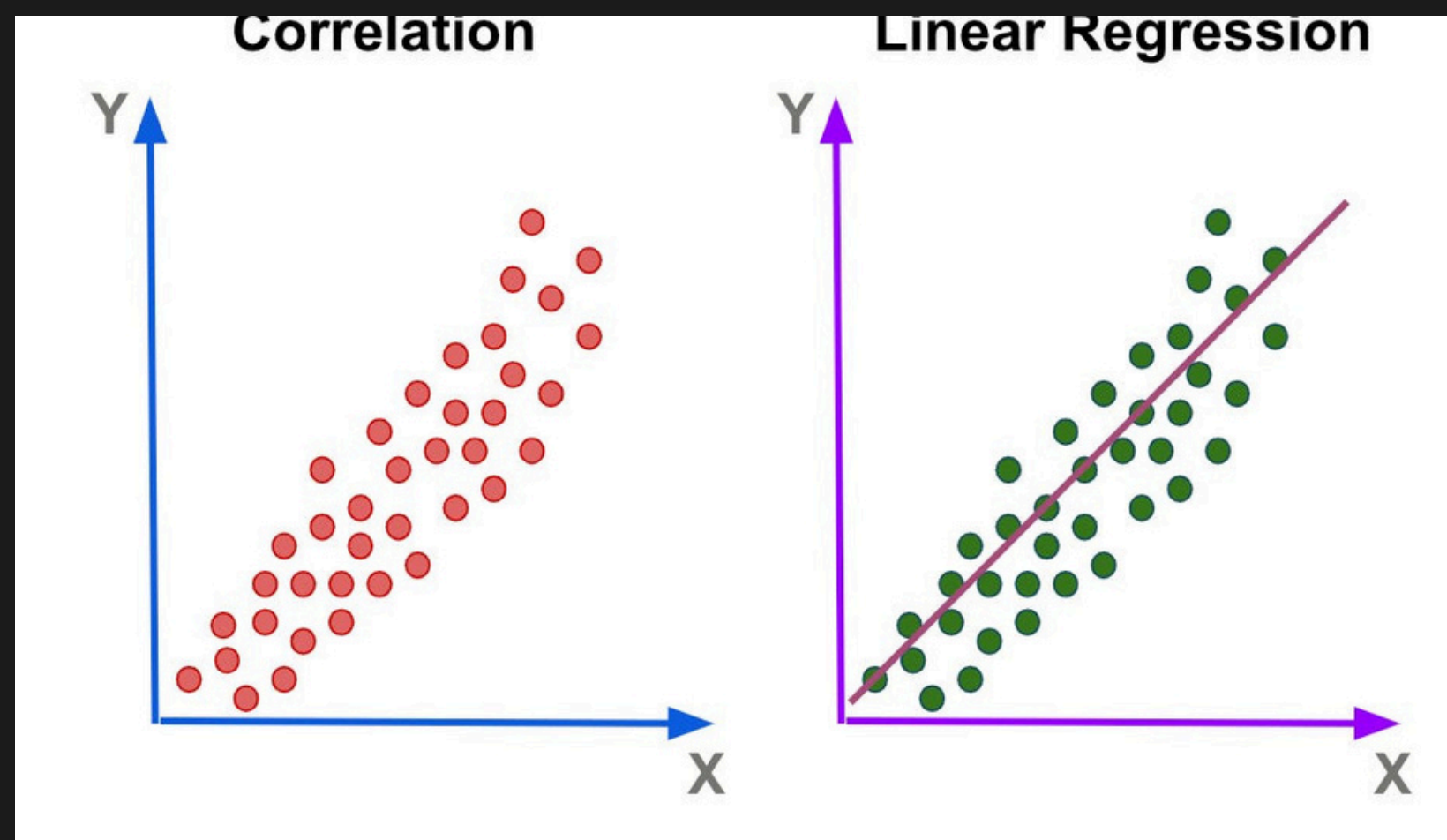
- Using the combined feature set, I want to apply clustering algorithms K-means or hierarchical clustering to segment the data into meaningful groups.
- This segmentation is based on the similarity of features within the data.
- Also, we can determine the optimal number of clusters using methods like the Elbow method.



05 Cluster-Specific Regression Models

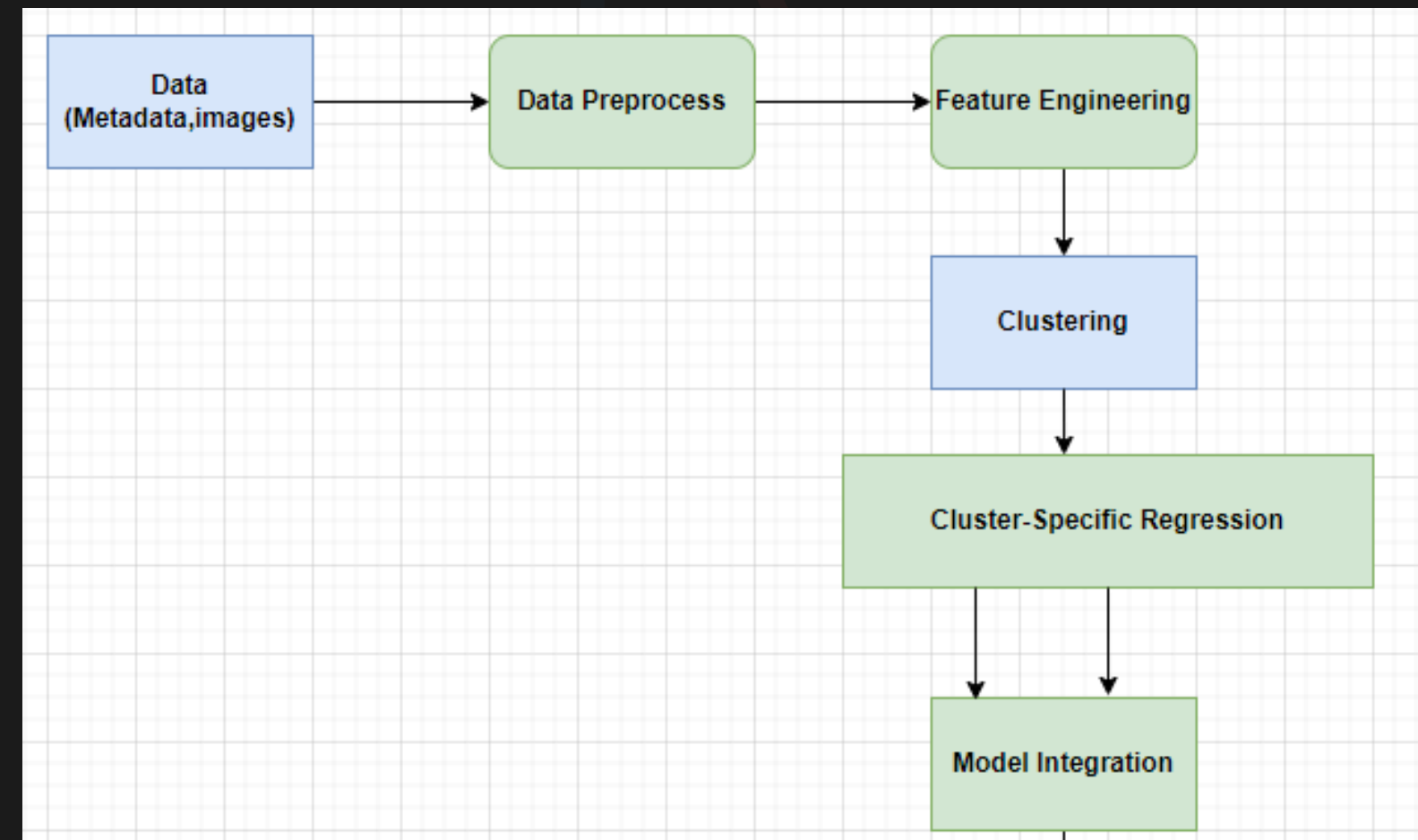


- For each cluster, we develop tailored regression models.
- Depending on the data characteristics of each cluster, we might use models ranging from Linear Regression to more complex ones.
- This allows us to specifically address the unique aspects of different groups within our dataset.

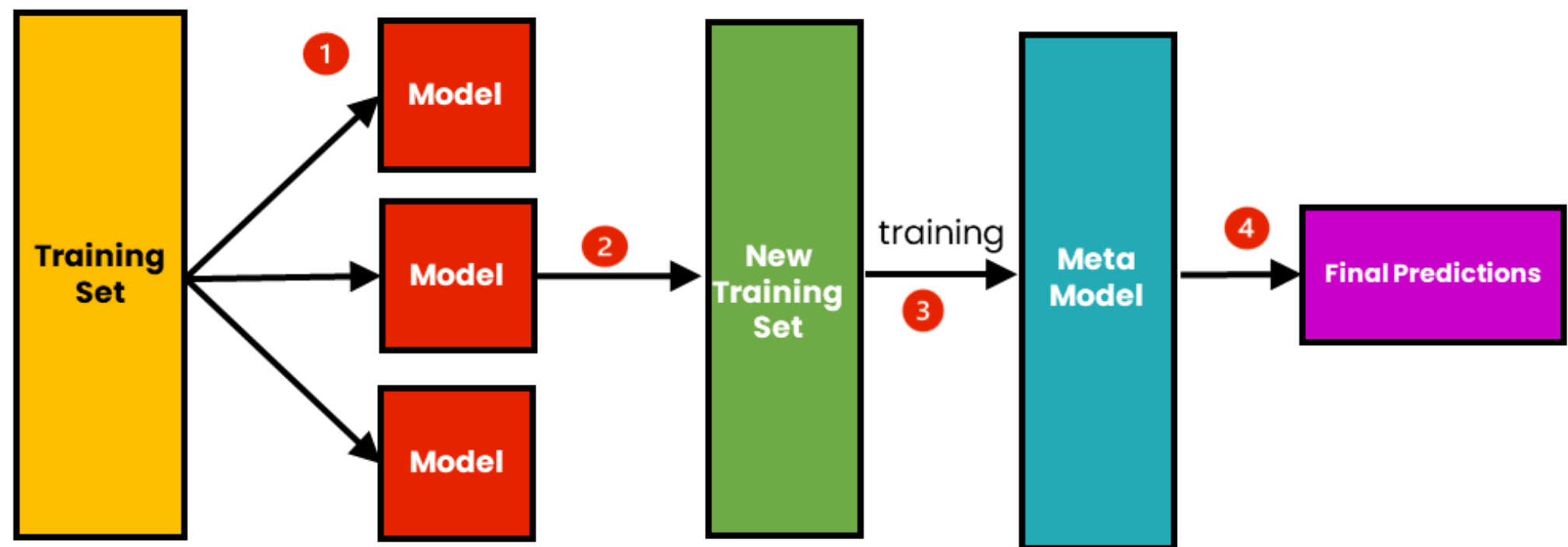


06 Model Integration and Ensemble

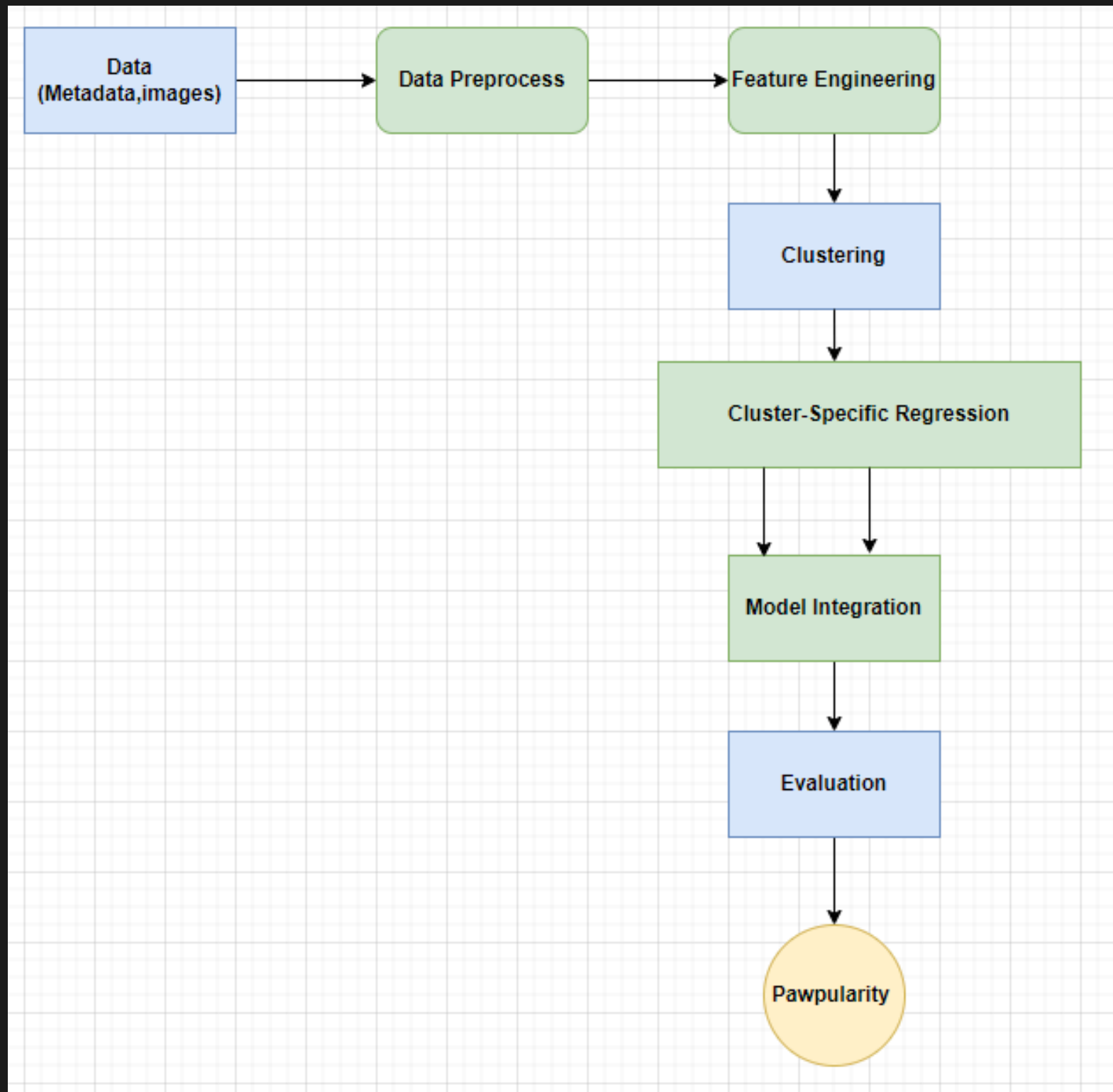
- After developing models for each cluster, we integrate their outputs using an ensemble method.
- One effective technique is stacking.
- Where a second-level model learns how to best combine predictions from all cluster-specific models, thereby enhancing overall prediction accuracy.



The Process of Stacking



06 Evaluate and Deploy

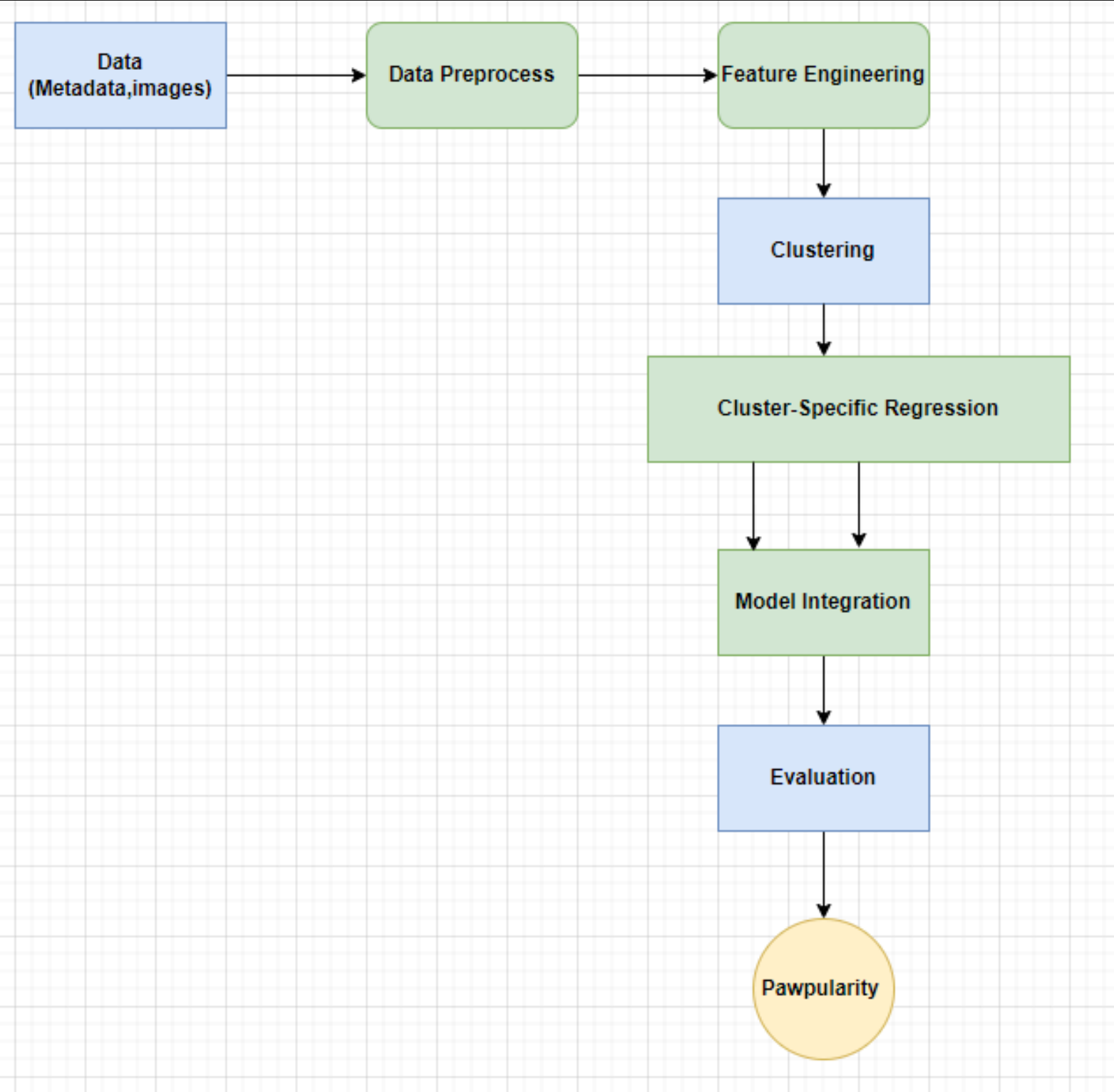


$$RMSE = \sqrt{\frac{\sum_{n=1}^N (\hat{y}_n - y_n)^2}{N}}$$

\hat{y}_n : Predictions
 y_n : Label

Evaluation: Root Mean Square Error (RMSE)

FLOW CHART



Finish by Date	Name
Aug 2	Feature Engineering
Aug 2	Clustering
Aug 9	Cluster-Specific Regression
Aug 16	Model Integratoion
Aug 18	Evaluate and Deploy

Final Date: 8/19

The background is a dark charcoal grey. It is filled with various abstract, hand-drawn style elements in bright colors: yellow, light blue, purple, orange, green, and pink. These include loops, swirls, straight lines, and geometric shapes like rectangles and circles. Some shapes contain horizontal lines, resembling simplified icons for a menu or a list. The overall effect is a busy, celebratory, and modern aesthetic.

**THANK YOU
FOR
LISTENING!**