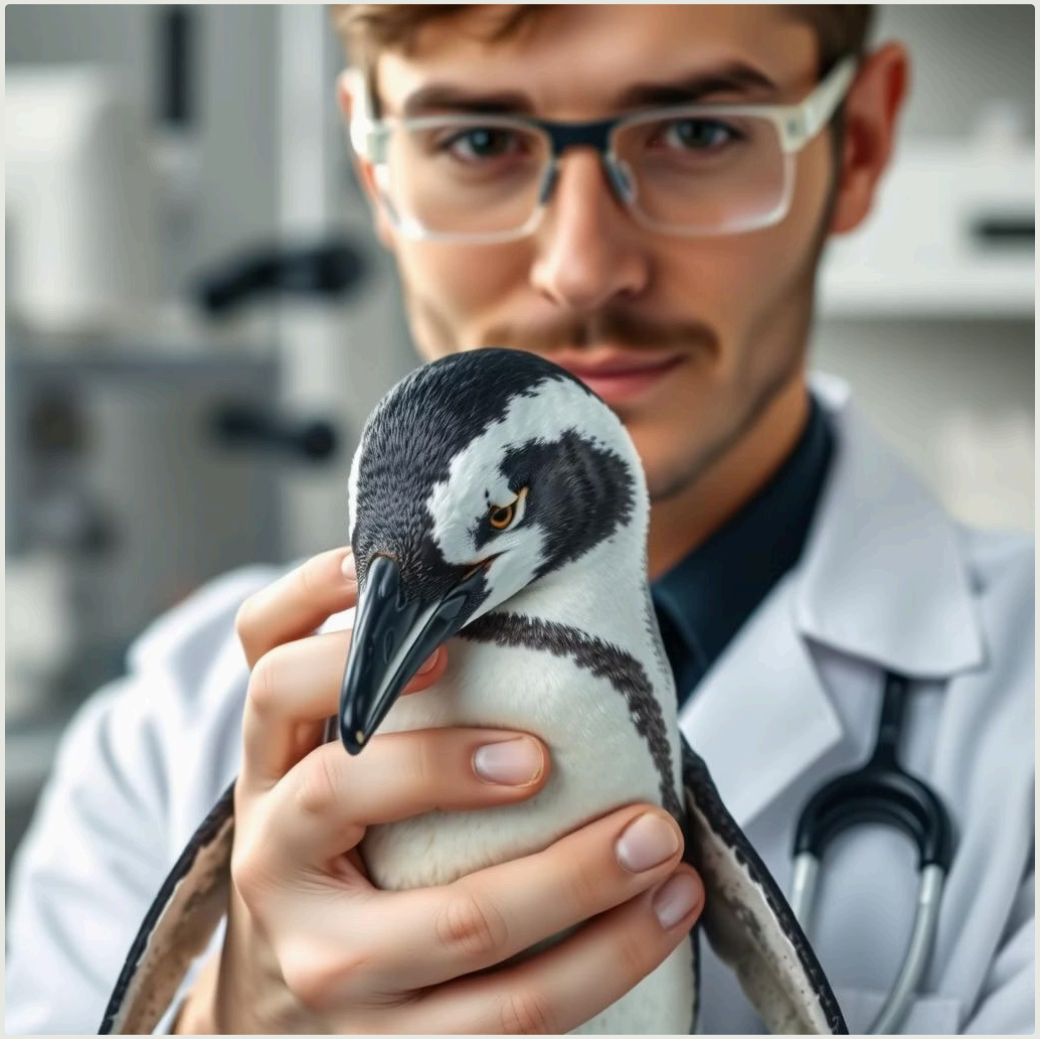
PENGUIN CLASSIFICATION USING CNN

Using the Palmer Penguins Dataset

A by Adon banker

THE PALMER PENGUINS DATASET



The Palmer Penguins dataset is a well-known resource for machine learning, providing detailed measurements for various penguin species.

- **344 samples**: Each representing an individual penguin.
- **7 features**: Including bill length, flipper length, body mass, sex, and island of origin.
- **3 target species**: Adelie, Chinstrap, and Gentoo.

PROBLEM STATEMENT: PREDICTING PENGUIN SPECIES

Our primary goal is to accurately predict the species of a penguin using its physical characteristics. This is achieved by leveraging deep learning techniques, specifically Convolutional Neural Networks (CNNs).

GOAL: PREDICT PENGUIN SPECIES

Utilize advanced deep learning models to classify penguin species based on their unique features.

METHOD: DEEP LEARNING

Employ a CNN architecture tailored for structured tabular data, demonstrating its versatility beyond image recognition.

CNN ARCHITECTURE FOR TABULAR DATA

1

CONV2D LAYERS

Initial layers to extract abstract features from the reshaped input.

2

MAXPOOLING

Reduces dimensionality, preserving essential information.

3

FLATTEN LAYER

Transforms 2D feature maps into a 1D vector for Dense layers.

4

DENSE LAYERS

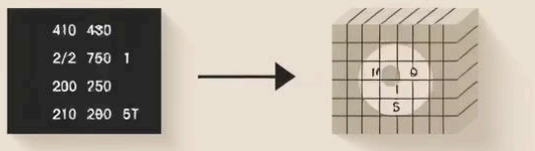
Fully connected layers for final classification.

5

SOFTMAX ACTIVATION

Outputs probabilities for each penguin species.

CREATIVE RESHAPING: TABULAR TO IMAGE-LIKE INPUT



A key innovation in this approach is the reshaping of tabular data into a format suitable for CNN processing.

INPUT RESHAPED TO (2, 3, 1)

Our 7 features are creatively arranged into a 2x3x1 "image" structure.

MIMICKING IMAGE DATA

This transformation allows the CNN to apply its powerful pattern recognition capabilities.

EXCEPTIONAL TRAINING RESULTS

The model achieved impressive results, demonstrating the effectiveness of CNNs on this unique tabular dataset.

98%

ACCURACY

Overall accuracy across all penguin species.

Penguin. Classion. Molul

Adelie	0	10	15	15
ADellie	1	12	20	20
Contre	9	10	20	15
Gentoo	11	13	15	10
Gentoo	13	21	15	40
	21	35	41	45
	Hand	Recarl	Peairl	Pegjon

Precission:	Detal	Actell	Retall	Miglit
Presuine	75,	66,	139	415
Sand Mist	4.1.	48,	120	335
Pale	2.5	275	190	0.

WHY CNNS FOR TABULAR DATA?

CAPTURES LOCAL FEATURE PATTERNS

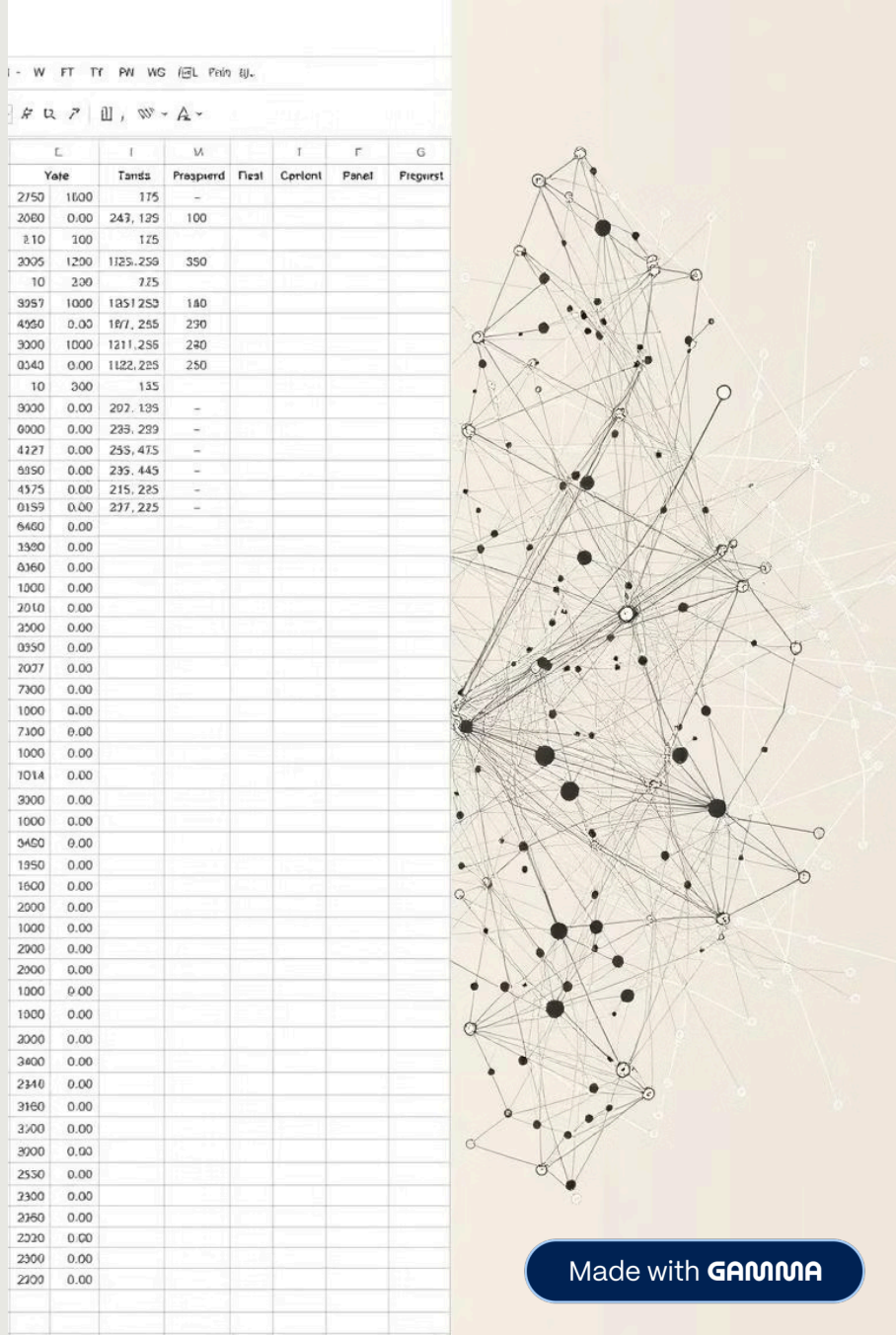
Even in reshaped tabular data, CNNs can identify subtle relationships between features.

CREATIVE RESHAPING BENEFITS

Transforms linear data into a grid, enabling convolutional filters to scan for patterns.

ROBUST FEATURE EXTRACTION

CNNs excel at automatically learning relevant features, reducing the need for manual feature engineering.



CONCLUSION AND FUTURE POTENTIAL

CNNS BEYOND IMAGES

This study demonstrates the powerful applicability of CNNs to non-image, tabular datasets.

ACCURATE & FLEXIBLE MODEL

The resulting model is highly accurate and adaptable for various classification tasks.

DEEPER TUNING POTENTIAL

Further optimization of hyperparameters and architectures could yield even higher performance.