Feature Engineering and Selection

Dr. Tassadaq Hussain
Assistant Professor Riphah International University

Collaborations:
Microsoft Research and Barcelona Supercomputing Center
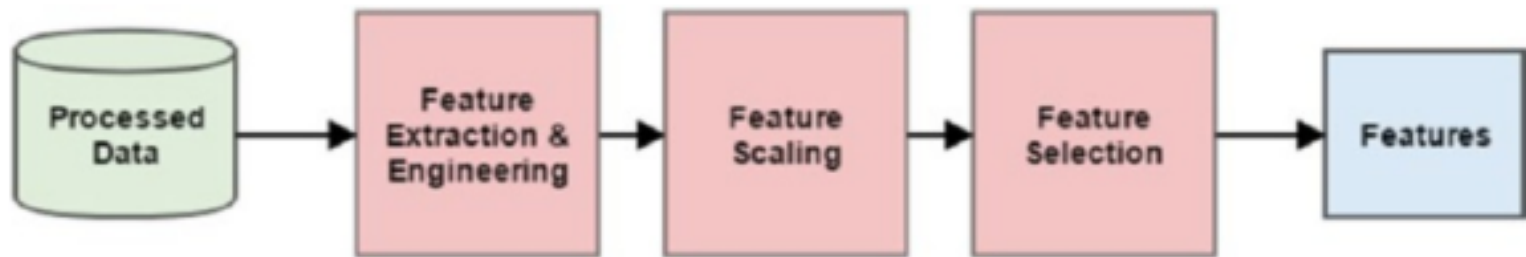### Barcelona, Spain
UCERD Pvt Ltd Islamabad

- Feature extraction and engineering
- Feature scaling
- Feature selection

UCERD

Features: Understand Your Data Better

- Data and Datasets

- Features

- Models

-

Feature Engineering

"Coming up with features is difficult, time-consuming, requires expert knowledge. 'Applied Machine Learning' is basically feature engineering."

—Prof. Andrew Ng

"Feature engineering is the process of transforming **raw data** into **features** that better represent **the underlying problem** to **the predictive models**, resulting in improved **model accuracy** on **unseen data**."

—Dr. Jason Brownlee

"At the end of the day, some Machine Learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used."

—Prof. Pedro Domingos

- Raw data: This is data in its native form after data retrieval from source. Typically some amount of data processing and wrangling is done before the actual process of feature engineering.

- Features: These are specific representations obtained from the raw data after the process of feature engineering.

- The underlying problem: This refers to the specific business problem or use- case we want to solve with the help of Machine Learning. The business problem is typically converted into a Machine Learning task.

- The predictive models: Typically feature engineering is used for extracting features to build Machine Learning models that learn about the data and the problem to be solved from these features. Supervised predictive models are widely used for solving diverse problems.

- Model accuracy: This refers to model performance metrics that are used to evaluate the model.

- Unseen data: This is basically new data that was not used previously to build or train the model. The model is expected to learn and generalize well for unseen data based on good quality features.

UCERD

Why Feature Engineering?

- Better representation of data:

- Better performing models:

- Essential for model building and evaluation:

- More flexibility on data types:

- Emphasis on the business and domain:

UCERD

Features Engineering Data

- Numeric data

- Categorical data

- Text data

- Temporal data

- Image data
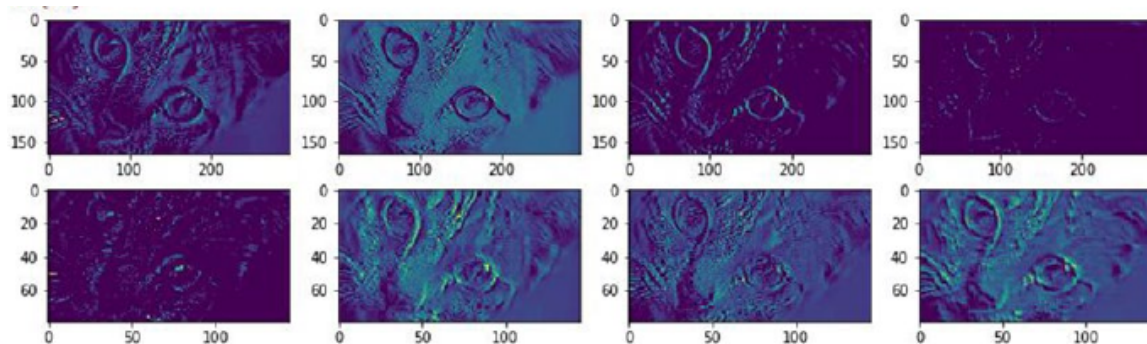
Intermediate feature maps obtained after passing though convolutional Layers