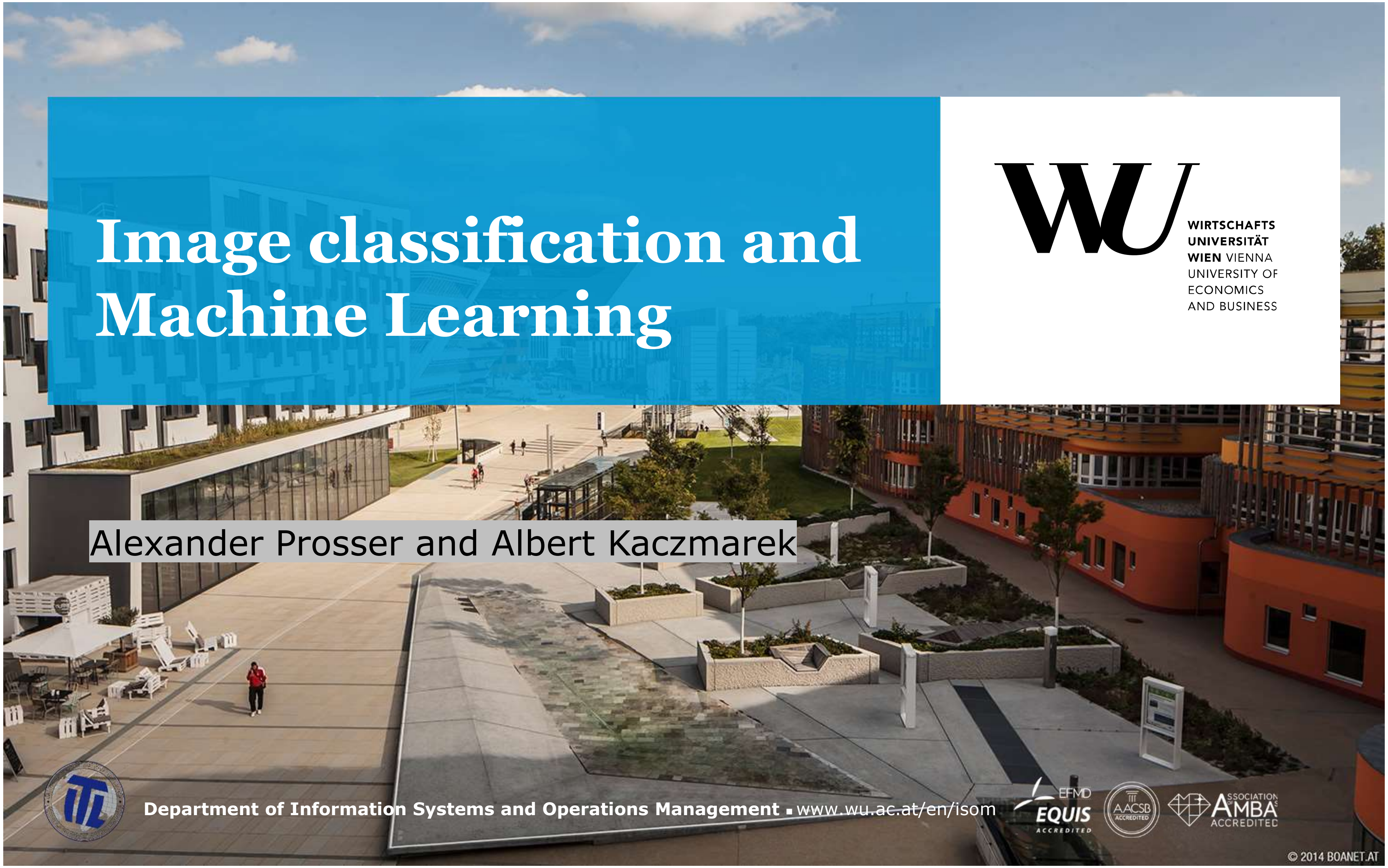
Image classification and Machine Learning

Alexander Prosser and Albert Kaczmarek



WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Department of Information Systems and Operations Management ▪ www.wu.ac.at/en/isom

The Case

A tire refurbishment company selects tires that are worth being selected for refurbishment. Thus far, the selection was done by visual inspection.

Now ML is to be applied.

There are a few thousand pre-classified tires to learn from and some also pre-classified tires as test sample.

Then operations start and the first batch of 1000 un-pre-classified tires has to be processed.

The Case

On top, the classification result

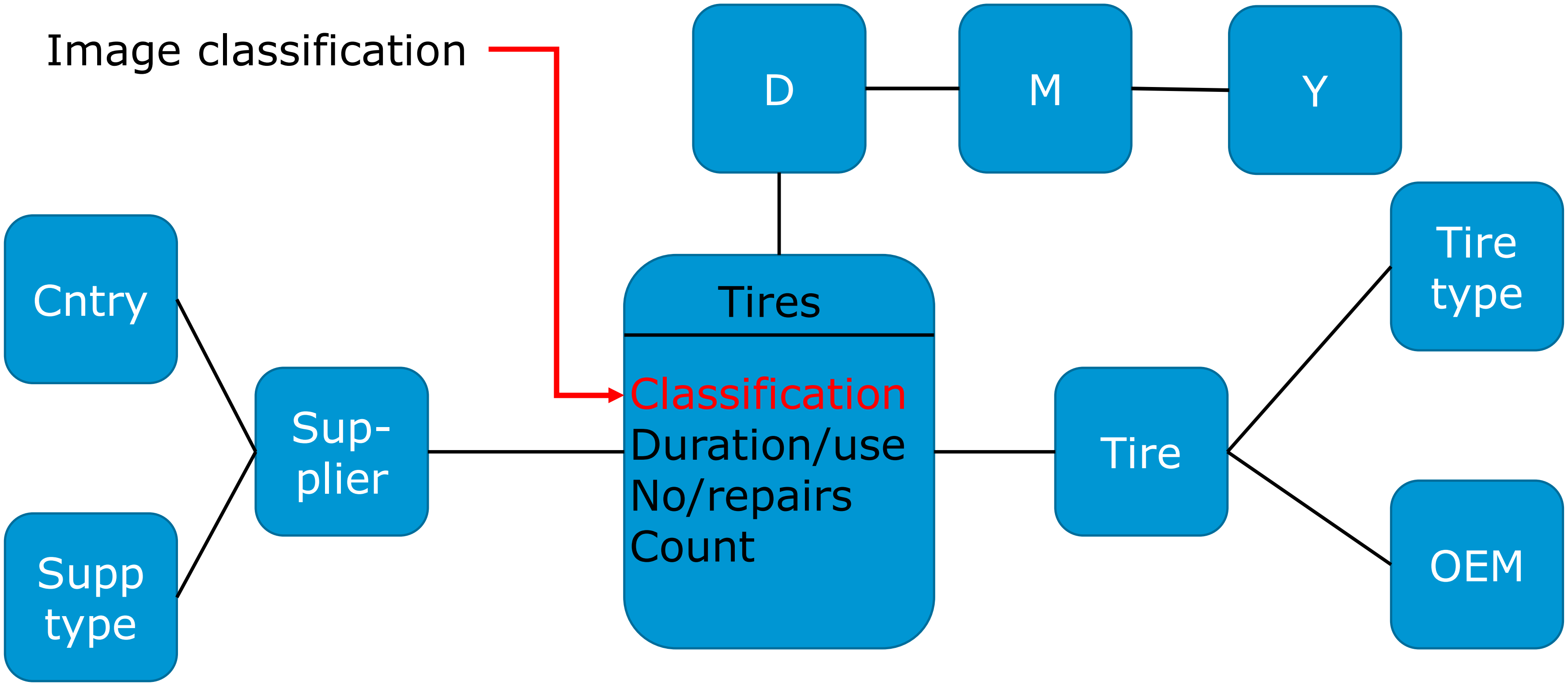
useful = positive

to be discarded = negative

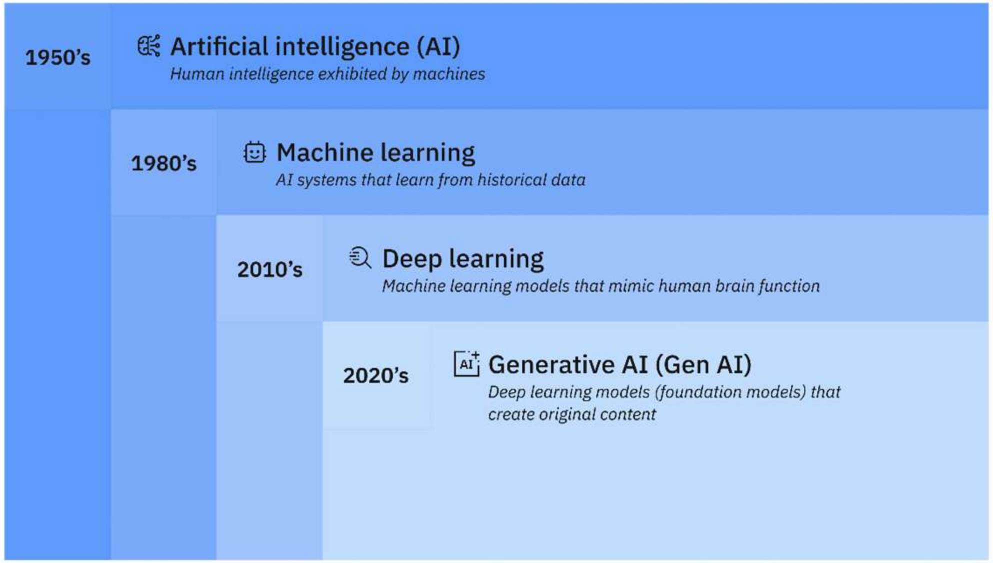
is to be analysed according to the Digital Product Passport.

This will be done in SAP BW/4Hana and follows the data model depicted below:

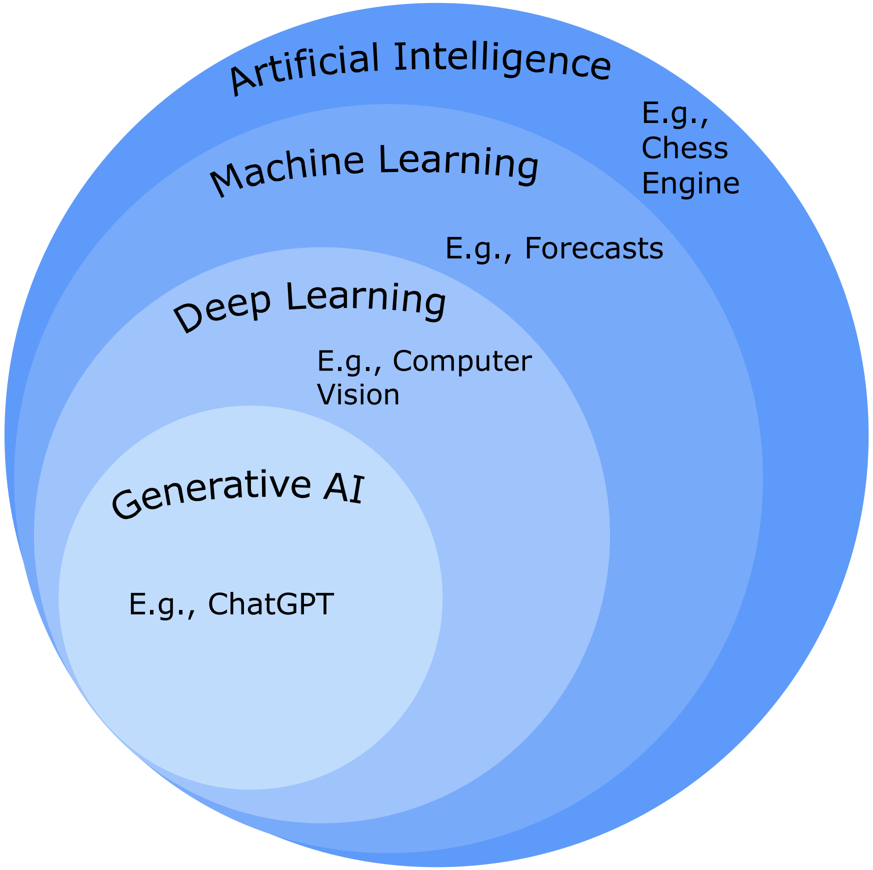
The Case – DFM of the digital passport data



Classification of AI



Types of Artificial Intelligence (IBM, 2024)



Supervised vs Unsupervised Learning

- **Supervised Learning** – a machine learning technique that uses human-labeled input and output datasets to train artificial intelligence models. The trained model learns the underlying relationships between inputs and outputs, enabling it to predict correct outputs based on new, unlabeled real-world input data. (IBM, 2024)

- **Unsupervised Learning** – uses machine learning (ML) algorithms to analyze and cluster unlabeled data sets. These algorithms discover hidden patterns or data groupings without the need for human intervention. (IBM, 2021)

- **Supervised Learning**

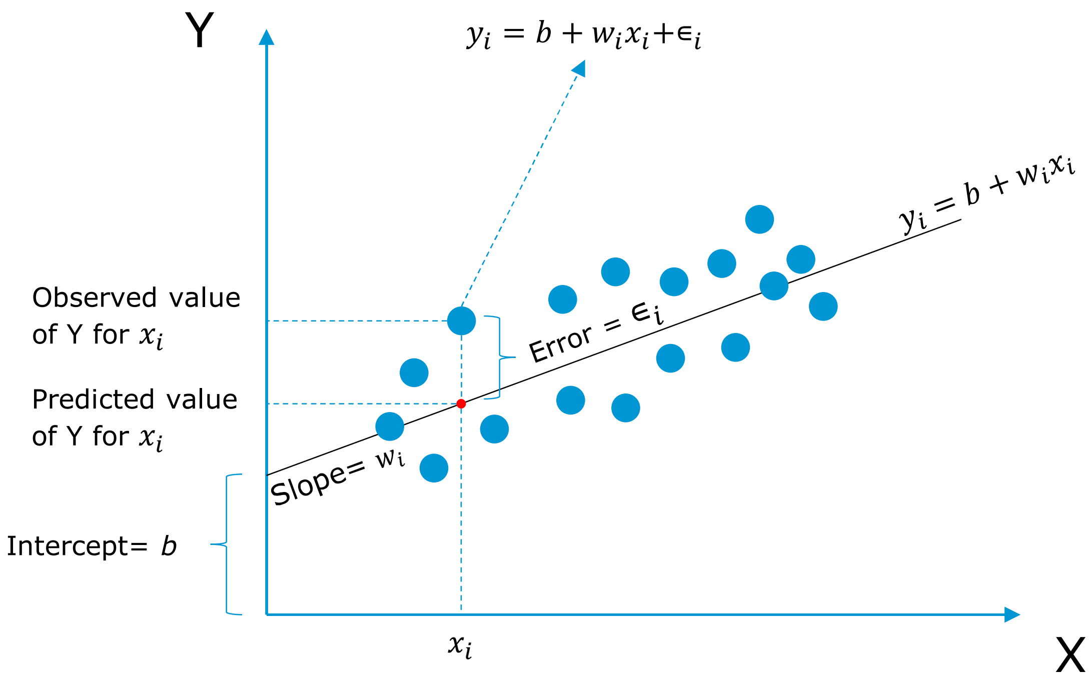
"A human provided me with labeled data: images categorized as either "Cat" or "Dog." I analyzed these labeled images and learned the distinguishing features of each class. Now, based on what I learned, I can classify new images as either a cat or a dog."

- **Unsupervised Learning**

"Without being given explicit labels, I analyzed many images along with associated words. I noticed that images featuring creatures with spiky ears and piercing eyes often appear with the word "Cat," while images with smiling creatures and floppy ears often appear with the word "Dog." From these patterns, I learned to distinguish cats from dogs."

Theoretical background

Parameters: (The predicted value
e.g., a predicted class)



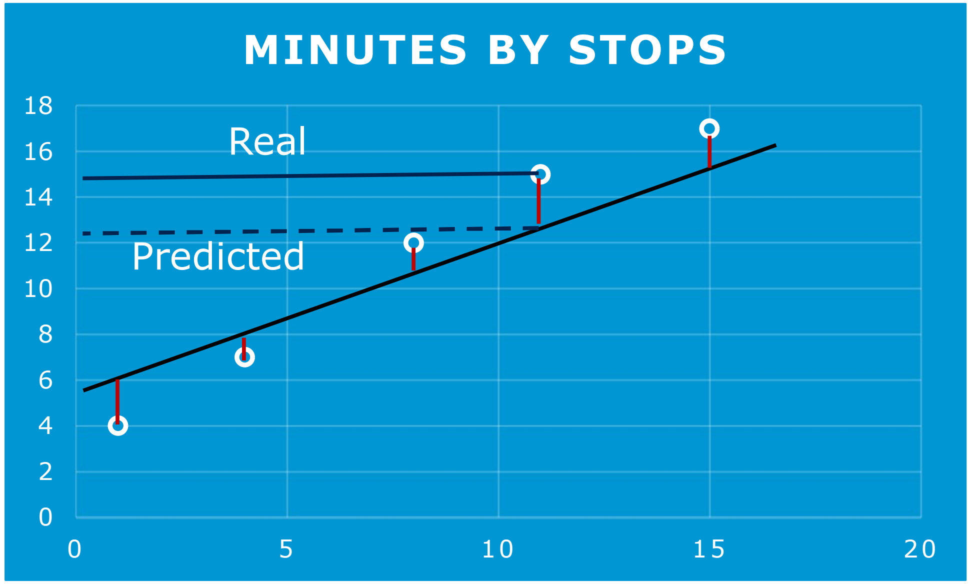
(The input variable e.g., a pixel)

w_i – weight
b – bias

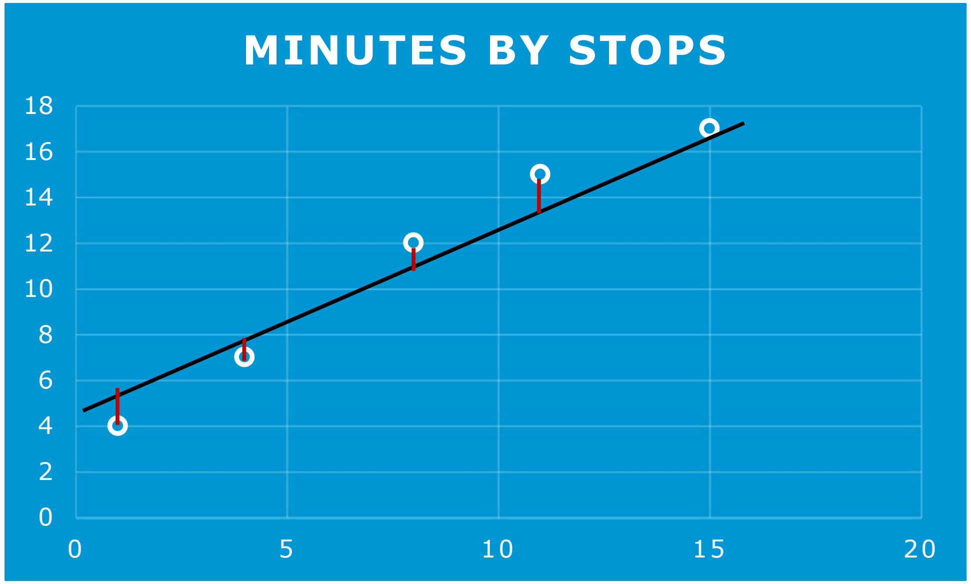
Gradient Descent, Simple Example

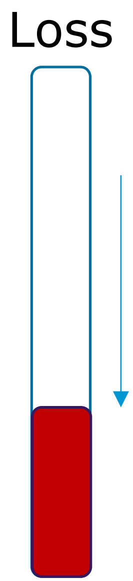
This is the relationship between traffic stops and the duration of the car travel to a book shop.

Stops	min
1	4
4	7
8	12
11	15
15	17



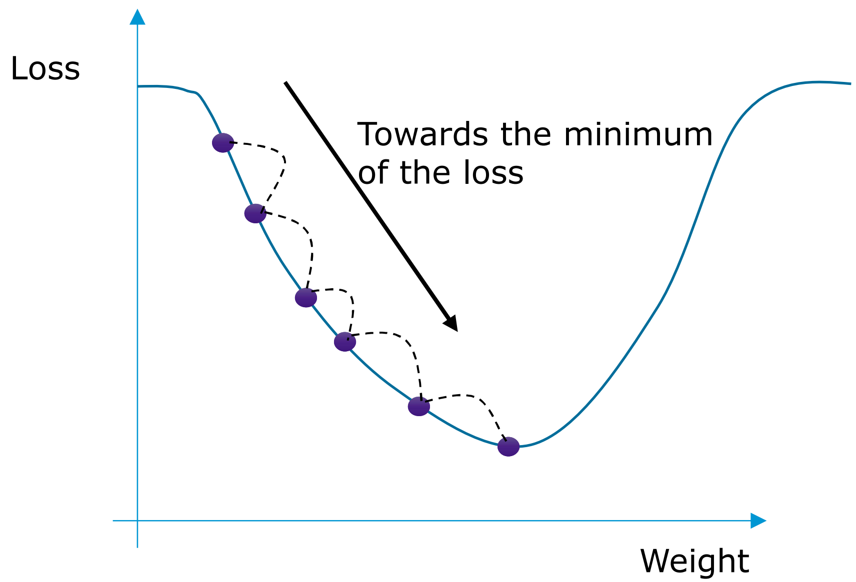
Gradient Descent, Simple Example





Adjust gradient
Weight to optimal
fit = minimal loss

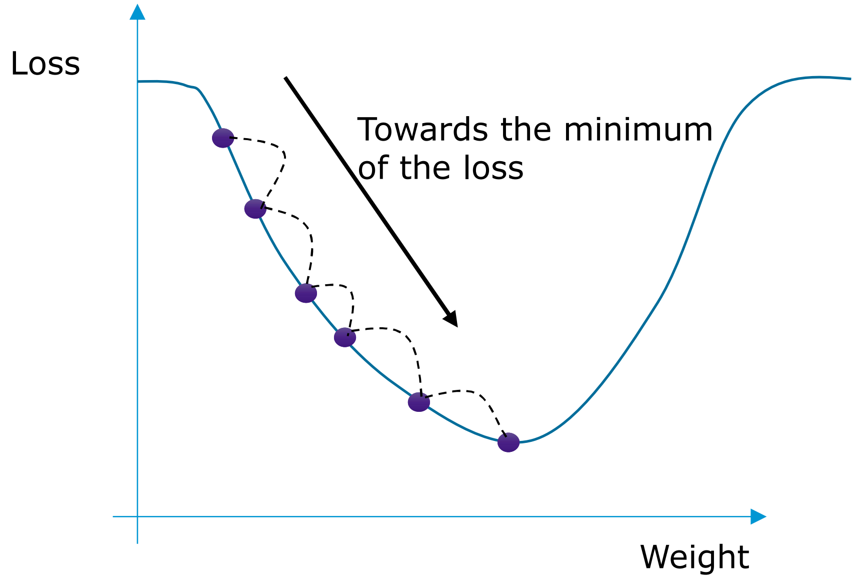
The Loss



The Loss

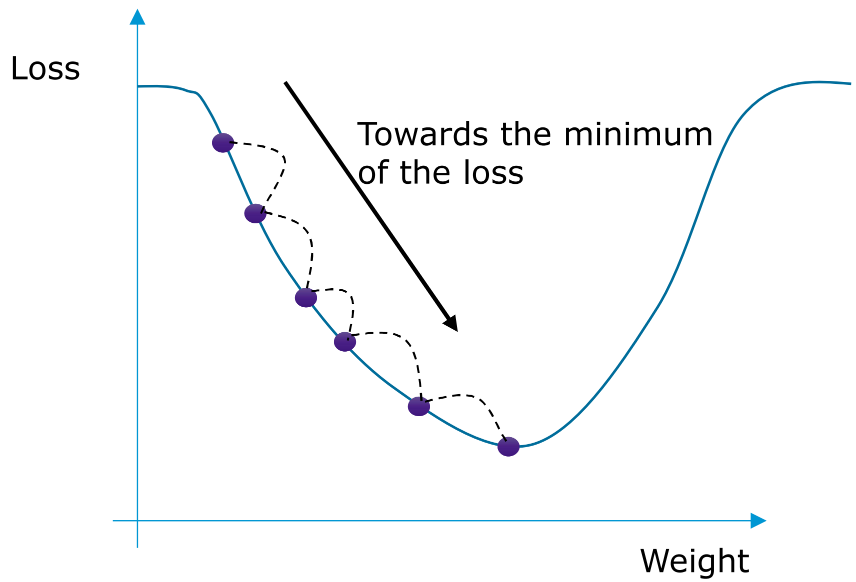
- **Learning Rate** – the step size taken into the gradient direction

- **Gradient-based optimization** – the method of optimizing the objective function
  - Batch gradient descent – iterates (recalculates the objective function) for all samples in dataset.
  - Mini-batch gradient descent – chooses a small group of data points per iteration.
  - Stochastic gradient descent (SGD) – randomly selects a data point per iteration.



The Loss

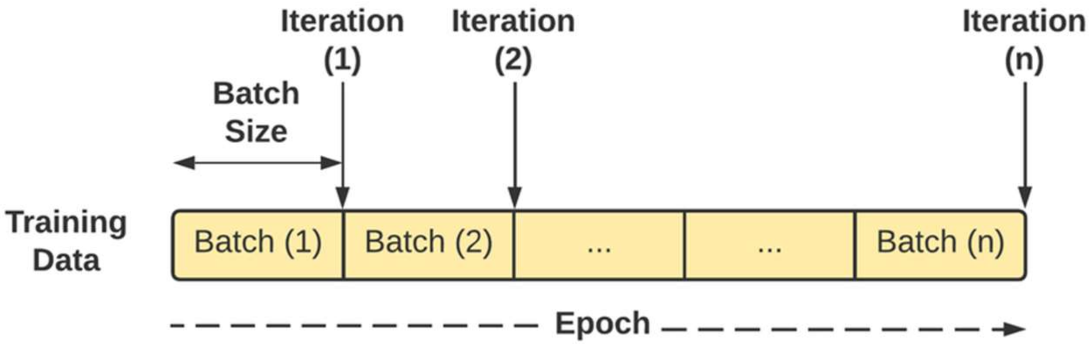
- Stochastic gradient descent (SGD) – randomly selects a data point per iteration.
  - Stochastic gradient descent with momentum (SGDM) remembers the update at the last iteration.
    Next update linear function of the gradient and the previous update.
    https://www.nature.com/articles/323533a0
  - RMSPROP (Root Mean Square Propagation), where learning rate is adapted for each parameter using running averages.
    RMSProp Definition | DeepAI
  - Adaptive Moment Estimation (ADAM) combining the above.
    [1412.6980] Adam: A Method for Stochastic Optimization



Theoretical background

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

- **Epoch** - It defines the number of times the entire data set has to be worked through the learning algorithm. Every sample in the training dataset has had a chance to update the internal model parameters once during an epoch. One or more batches make up an epoch. (Simplilearn.com, 2024)
- **Batch size** – "the number of images used in every epoch to train the network. Setting this hyperparameter too high can make the network take too long to achieve convergence (no more gain in accuracy); however, if it is too low, it will make the network bounce back and forth without achieving acceptable performance." (Kandel I., Caste K., 2020)



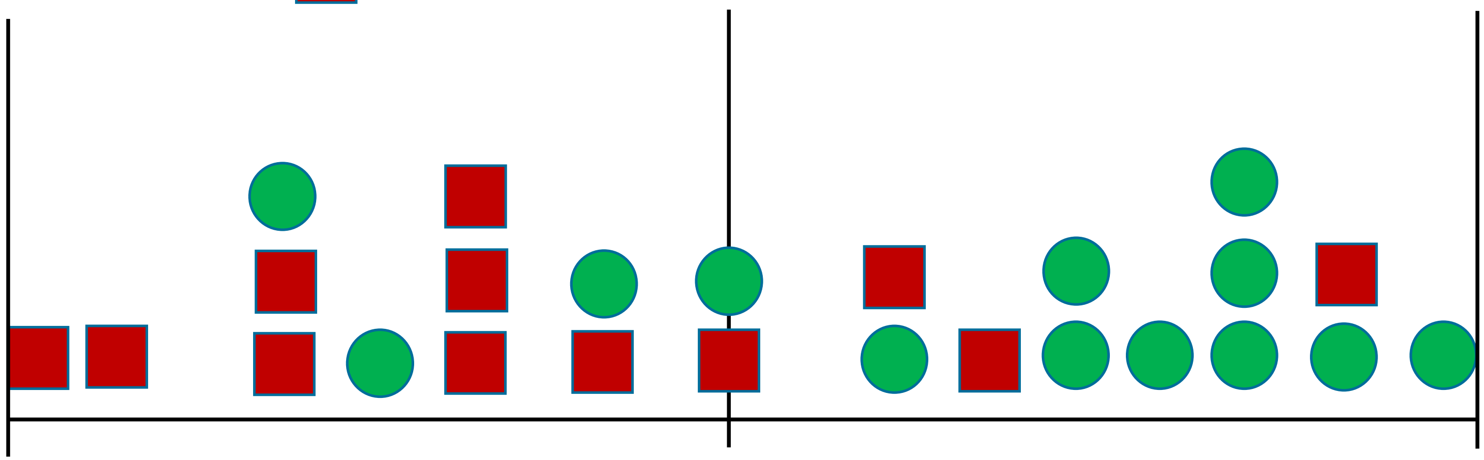
Source: Thanapol, P., et al., 2024

Confidence level in classification

100% confidence that case belongs to category

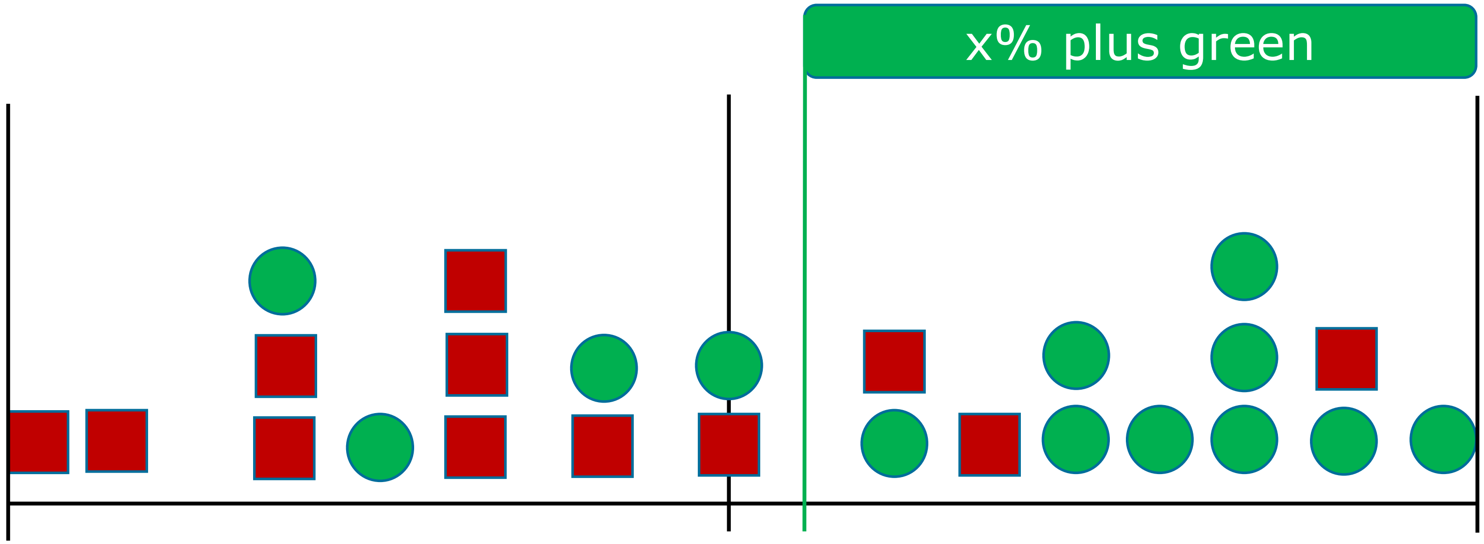
50% confidence that case belongs to either category

100% confidence that case belongs to category



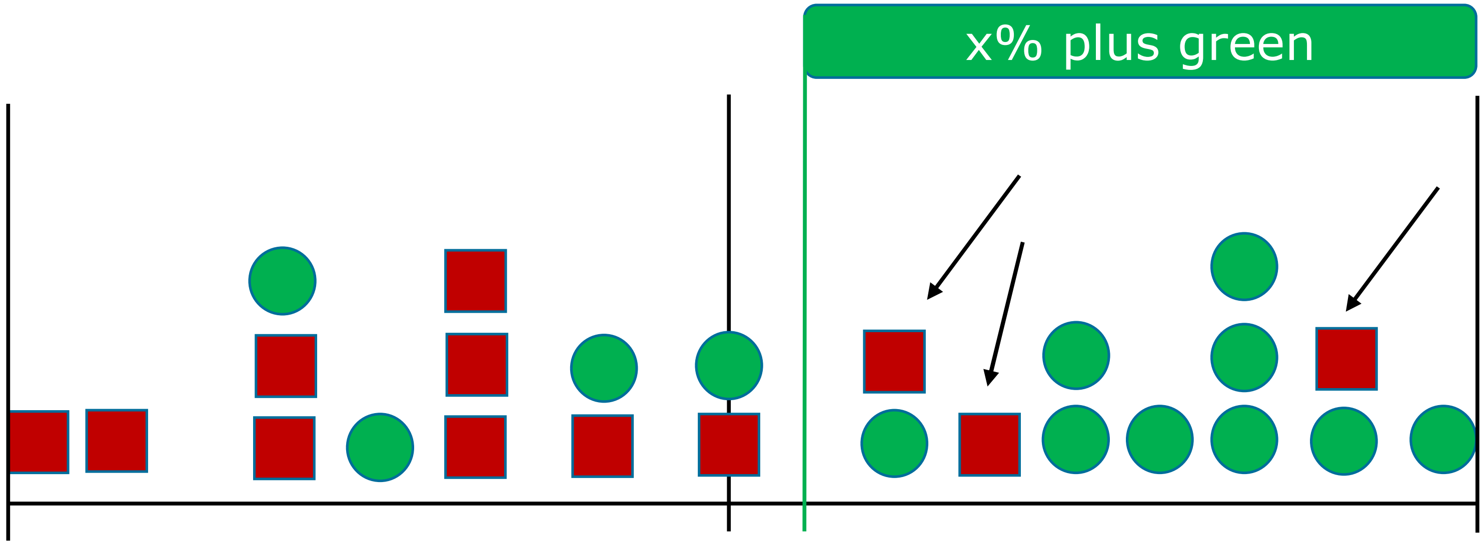
Confidence threshold

For green



Confidence threshold

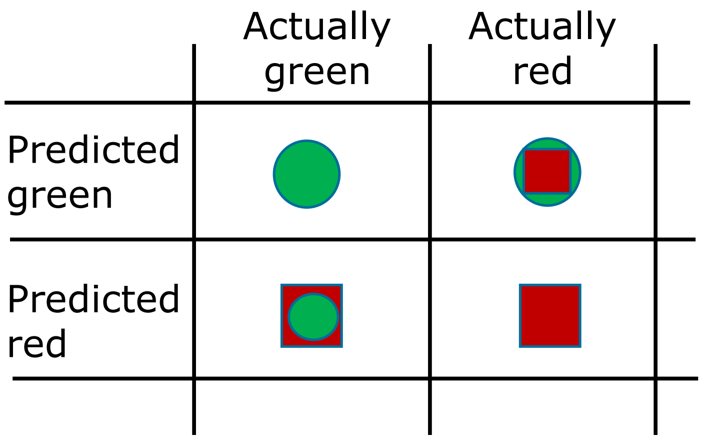
For green



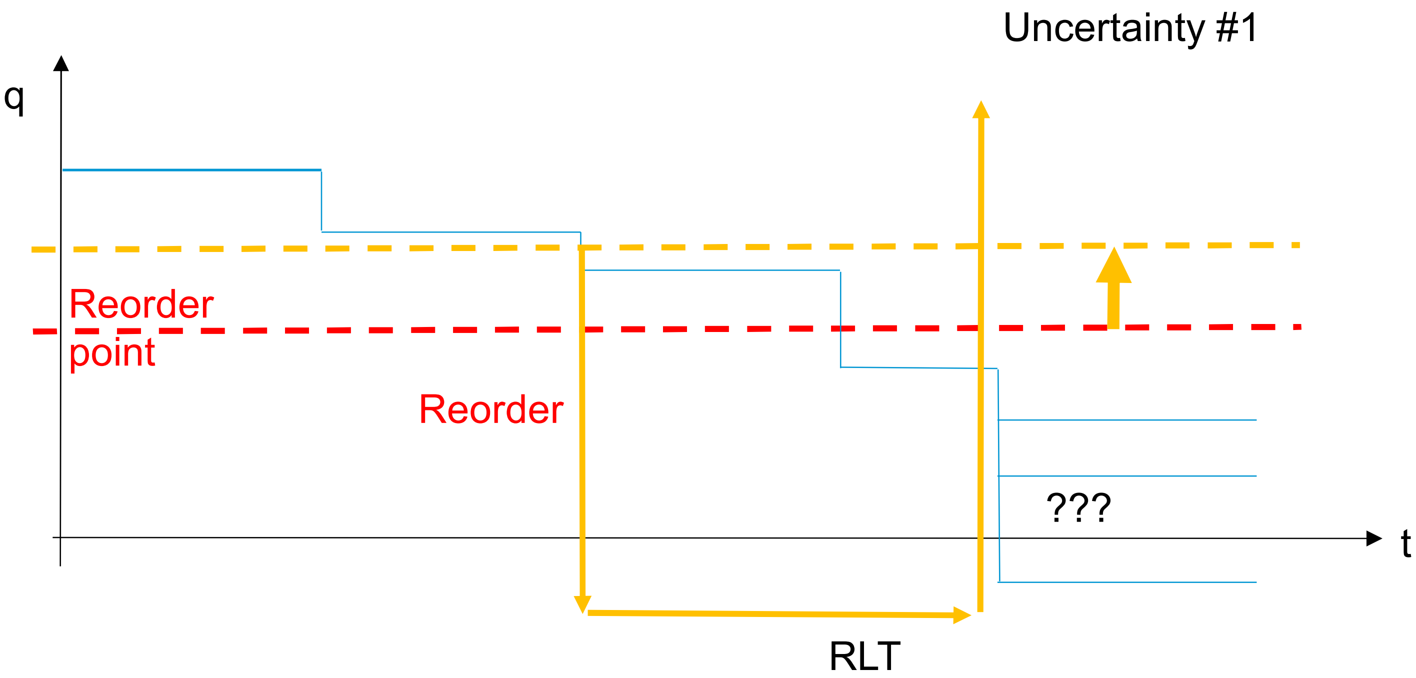
False positive vs false negative results

- **False positive** – negative-class instances incorrectly identified as positive cases (E.g., a bad tire classified as a good tire)
- **False negative** – the actual positive instances erroneously predicted negative (E.g., a good tire classified as a bad tire)

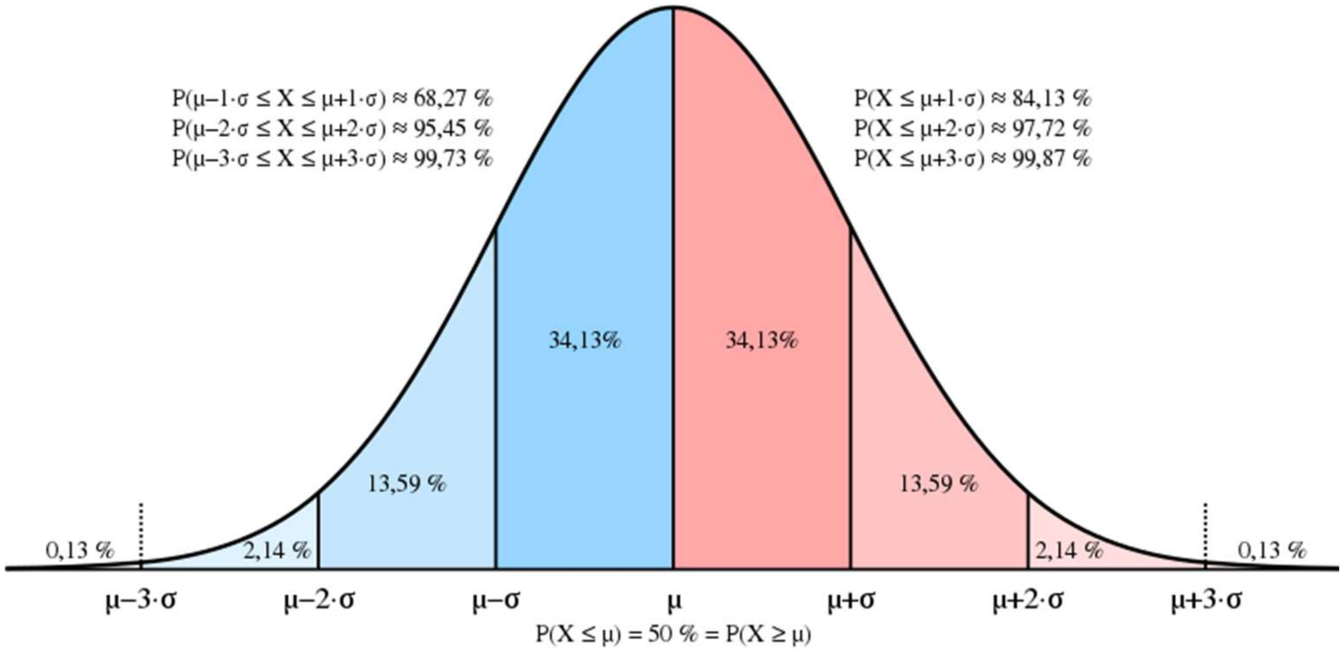
Confusion matrix



Consumption-based planning



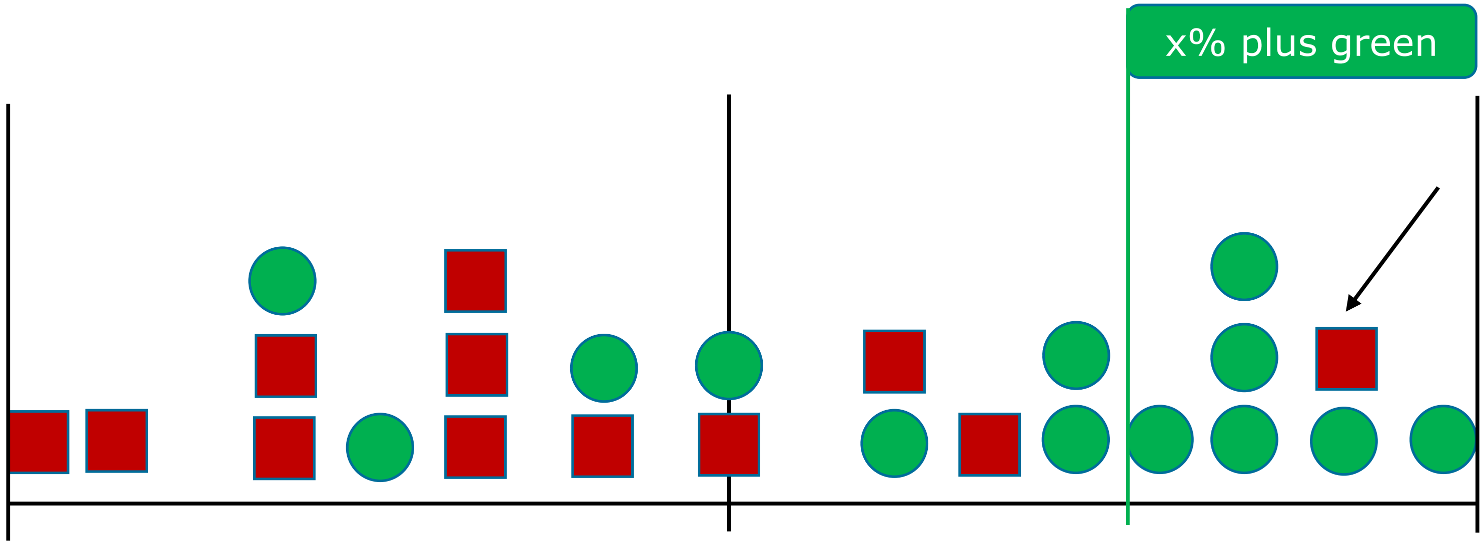
Consumption-based planning



Source: https://commons.wikimedia.org/wiki/File:Normal_Distribution_Sigma.svg, Creative Commons

Confidence threshold

For green



Model Evaluation Metrics

It depends on the application, which one has the higher negative impact.

Accuracy

- Calculates the percentage of correct responses in tasks such as classification or question-answering.

Recall

- Measures the actual number of true positives, or correct predictions, versus false ones in LLM responses.

F1 score

- Blends accuracy and recall into one metric. F1 scores range 0–1, with 1 signifying excellent recall and precision.

Prediction outcome

Example:

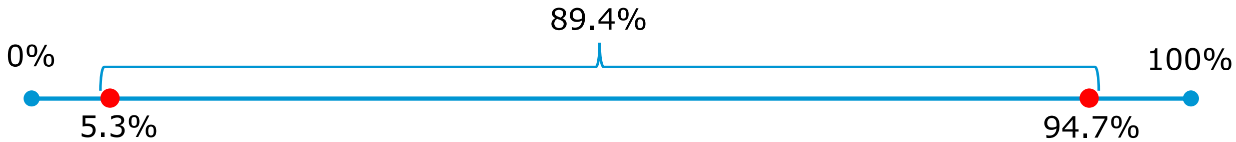
Image Name	Predicted Class	Prediction Probability	Confidence
123456.jpg	Normal	94.69523	89.39046

Probability

- Out of 2 classes, the model is 94.7% sure, that the tire is Normal/Good.

Confidence

- The model’s top prediction is 89.4 percentage points more likely than its nearest rival (second highest prediction).



Question to students: What is the confidence level, if prediction probability of one class is exactly 50%?

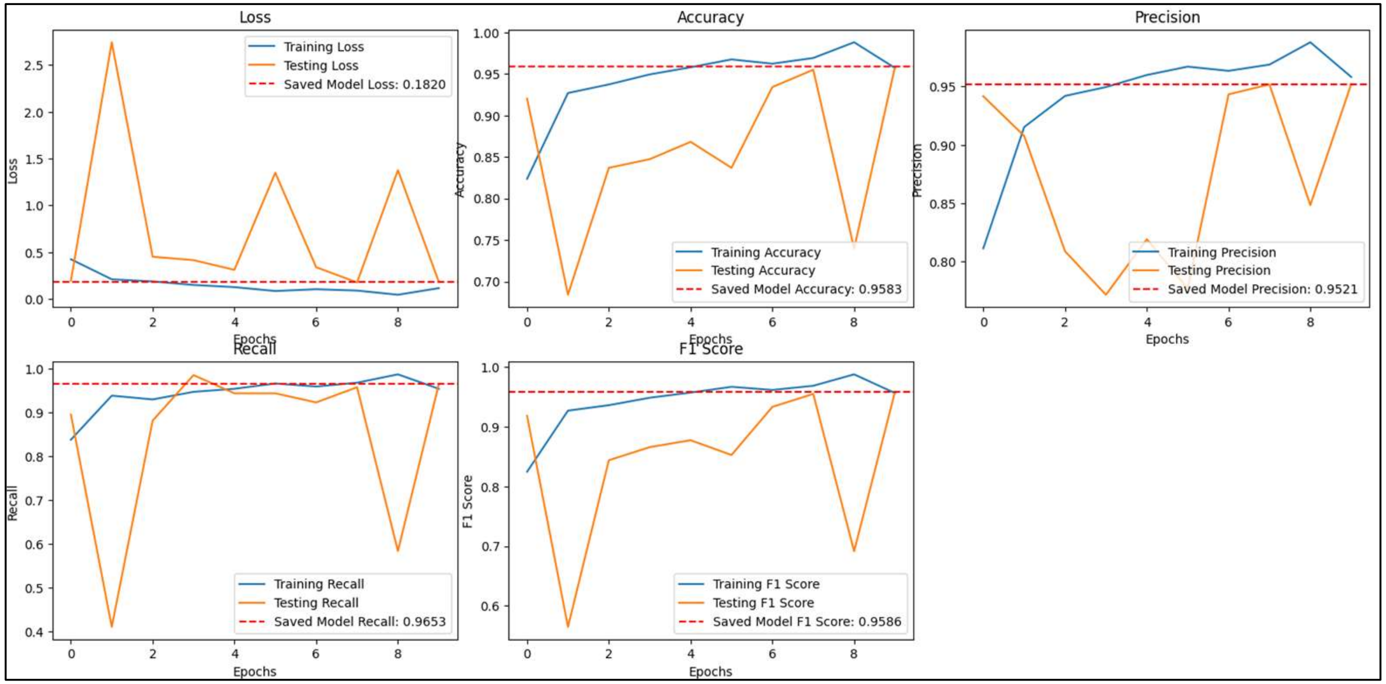
Answer: 0%

Using the model

After the training, you will get 2 files:

1. Tires_classification_model   5/29/2025 11:42 PM   PTH File   92,184 KB

2.

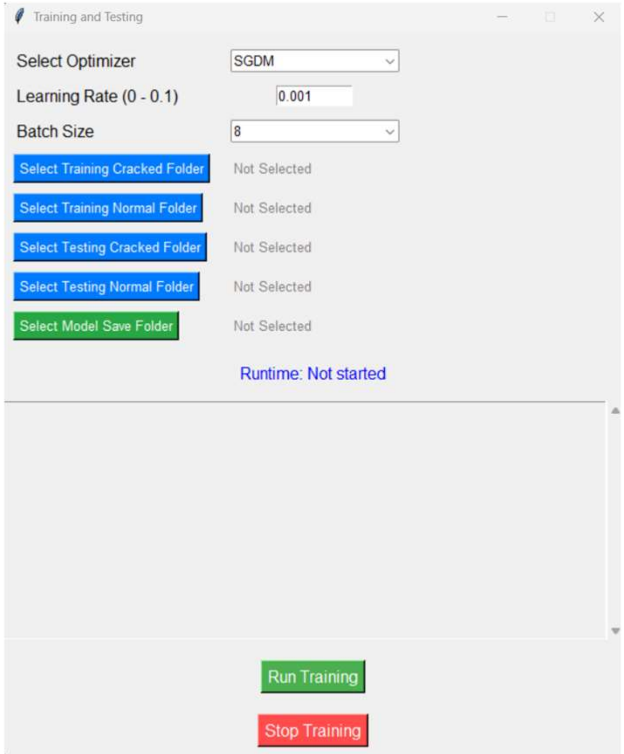


Setting up the training process

As mentioned earlier, finding the optimal set of hyperparameters is an experimental process. It typically involves selecting a combination of hyperparameters, running the training process, evaluating performance, adjusting the hyperparameters, and repeating this cycle until a high-accuracy model is achieved.
Fortunately, researchers have already conducted hyperparameter tuning on a similar dataset, which allows us to leverage their findings and save significant time.
Given that training on the university's hardware takes approximately 45 minutes per run, performing a full hyperparameter tuning process ourselves would be impractical.

Task for students: Go to https://doi.org/10.3390/s23042177 and go through the article. Based the findings of the authors explain what is the most optimal set of hyperparameters.
What aspects may influence a difference in performance of our model compared to the one presented by the authors?

Pretrained Network	Hyperparameter Configuration				Overall Accuracy (%)
	Train–Test Split Ratio	Optimizer/Solver Algorithm	Initial Learning Rate	Batch Size	
VGG-16	0.80:0.20	ADAM	0.0003	10	91.70
GoogLeNet	0.85:0.15	SGDM	0.0001	10	91.70
AlexNet	0.70:0.30	SGDM	0.001	32	87.50
ResNet-50	0.80:0.20	RMSPROP	0.0001	10	**93.80**



Answer: Number of images, quality of images, number of Epochs