Introduction to Machine Learning with Python

David Schaupp | WS2024

WHO AM I

Work:

- Company: Fivesquare
  Position: Senior Data Scientist
  Years: 2024 -

- Company: UAS St. Pölten
  Position: Part Time Lecturer ML
  Years: 2023 -

- Company: Umdasch Group Ventures
  Position: Software Engineer - AI
  Years: 2018 - 2024

- Company: Doka
  Position: Production Engineer
  Years: 2015 - 2018

Education:

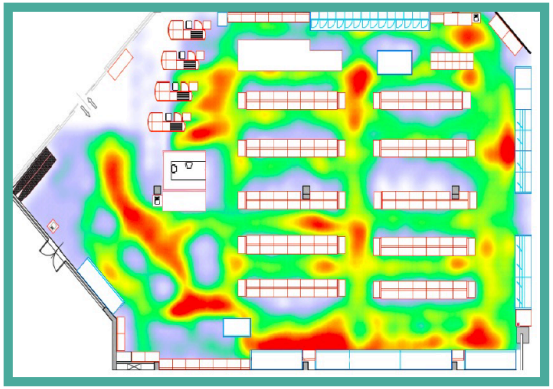
- University: UAS St. Pölten
  Study Programme: Smart Engineering
  Degree: Bachelor

- University: UAS St. Pölten
  Study Programme: Interactive Technologies
  Degree: Masters

- University: JKU Linz
  Study Programme: Artificial Intelligence
  Degree: Masters (on hold)

RL Machine Learning Projects

```
Brainstore:
    - Track people across multiple non-overlapping cameras
```



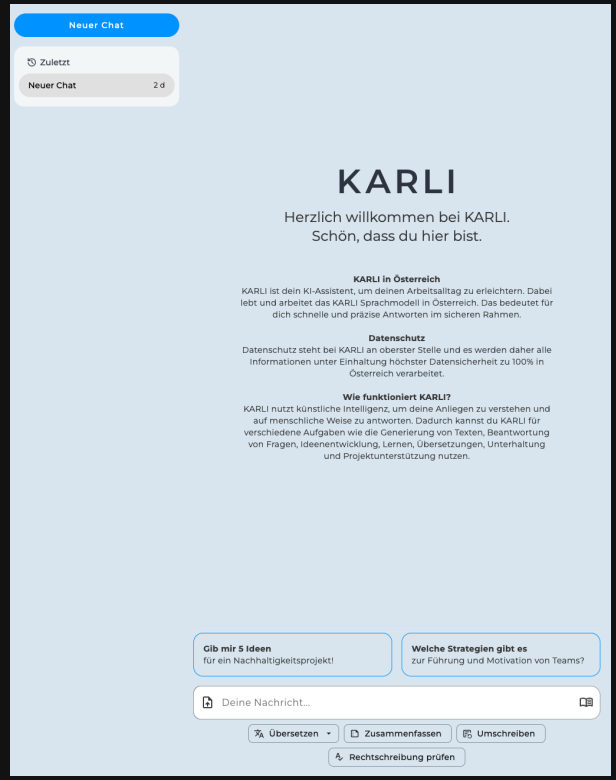
Tracking & Positional data → Movement & Customer journey → Visualization & Analytics

RL Machine Learning Projects

```
Predictive Maintenance:
    - Give an alarm when a machine is about to fail
```

RL Machine Learning Projects

```
KARLI LLM Assistant:
    - DSGVO compliant LLM Assistant hosted in Austria
```



Who are you?

```
Short Intro:
    - Name
    - Touchpoints with Python and/or Machine Learning
```

Course Procedure

```
10 appointments (3h each)

  - 24.09: 16:20-18:45
  - 01.10: 15:30-17:55
  - 08.10: 15:30-17:55
  - 29.10: 16:20-18:45
  - 05.11: 16:20-18:45
  - 12.11: 16:20-18:45
  - 19.11: 16:20-18:45
  - 03.12: 16:20-18:45
  - 10.12: 16:20-18:45
  - 17.12: 16:20-18:45
```

Grading 1/2

```
Mandatory Attendance:
    - 75% of the appointments
```

```
Assignments:
  - Bi-weekly assignments (5 in total)
  - 20 points per assignment
  - Group of 2 students
  - The solution will be presented by a randomly
    selected group in the following lecture
```

```
Appendix to Assignments:
  - Wrong submissions name leads to -5 points
  - f.e. "group_X_excercise_X.ipynb"
  - Plagiarism leads to 0 points
  - Using ChatGPT is not plagiarism as long
    as you can explain the code ;-)
```

Grading 2/2

```
Grading key assessment:
  - 100 .. 91 points - 1
  - 90 .. 81 points - 2
  - 80 .. 66 points - 3
  - 65 .. 51 points - 4
  - 50 .. 0 points - 5
```

Questions

```
Regarding Lectures and Content:
- eCampus
Personal Questions (f.e. Absence, ...):
- mail to: lbschauppd@fhstp.ac.at
- Teams
```

Quiz-Time

Content

```
Introduction to Machine Learning:
  - What is Machine Learning?
  - Machine Learning Project Checklist
Intro´s:
  - Colab
  - Python
  - Numpy
  - Matplotlib
Supervised Learning:
  - Classification (Binary|Multiclass)
  - Regression
  - Support Vector Machines
  - Decision Trees (Random Forest)
Unsupervised Learning:
  - Dimensionality Reduction
  - Clustering (k-means, DBSCAN)
Performances Measures:
  - Accuracy
  - Precision
  - Recall
  - F1-Score
  - ROC Curve
  - Confusion Matrix
```

Content

```
Introduction to Machine Learning:
  - What is Machine Learning?
  - Machine Learning Project Checklist
Intro´s:
  - Colab
  - Python
  - Numpy
  - Matplotlib
Supervised Learning:
  - Classification (Binary|Multiclass)
  - Regression
  - Support Vector Machines
  - Decision Trees (Random Forest)
Unsupervised Learning:
  - Dimensionality Reduction
  - Clustering (k-means, DBSCAN)
Performances Measures:
  - Accuracy
  - Precision
  - Recall
  - F1-Score
  - ROC Curve
  - Confusion Matrix
```

What is Machine Learning?

Arthur Samuel (1959):
- Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.

Easy terms:
- Machine Learning is just Maths

Matrix A × Matrix B = Magic! ✨

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

"Intelligence" = Math + More Math + Even More Math

Spam Filter Example

- Machine learning program that learns to flag spam emails automatically

Basic Jargon:

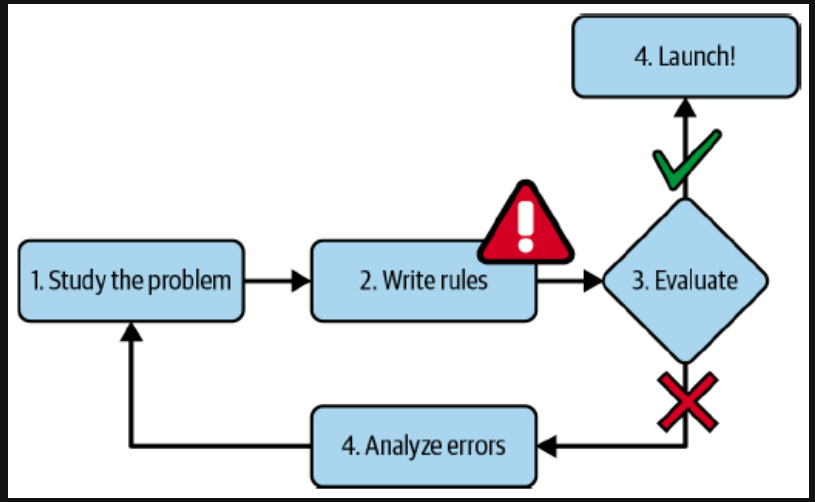
- The examples that the system uses to learn are called the training set

- Each training example is called a training instance (or sample)

- The part of a machine learning system that learns from data is called a model

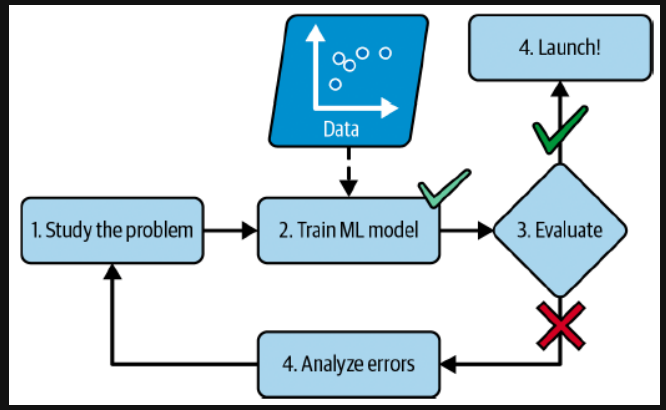
Spam Filter - Traditional Approach

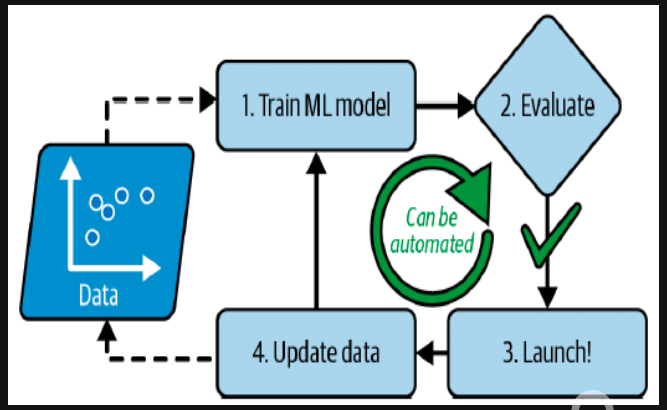
- 1: Examine what spam typically looks like ("credit card", "free", "buy now", ...)
- 2: Write detection algorithm with many rules ("credit card" in subject and "buy now" in body)
- 3: Test your program, repeat steps 1 and 2 until it is good enough
- 4: Launch the difficult program which will become a long list of complex rules that are hard to maintain
- Spammers notice that "buy now" is often flagged as spam and start writing "b.u.y. n.o.w." instead - endless loop



Spam Filter - Machine Learning Approach

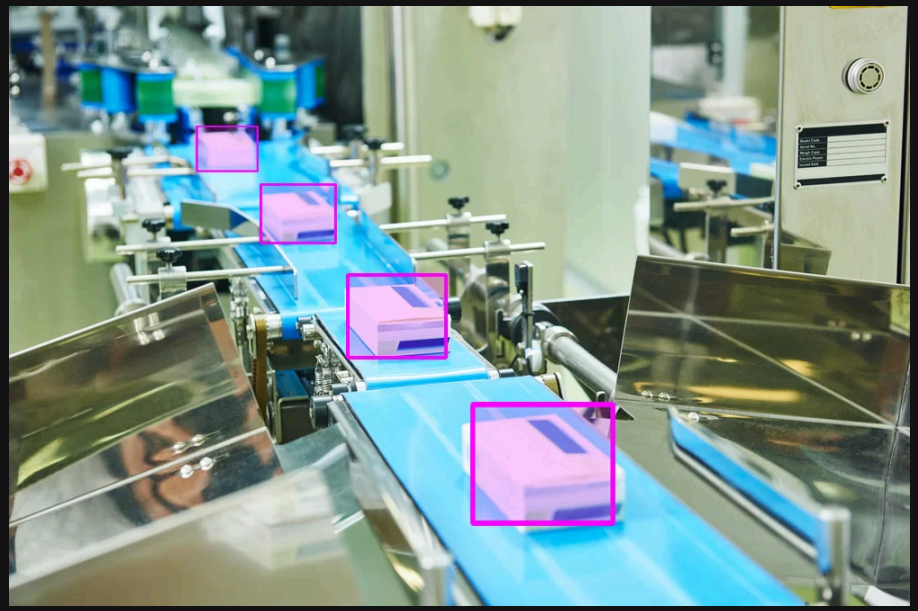
- 1: Collect many example emails that are correctly classified as spam or not spam (by humans)
- 2: Train a machine learning model to learn to classify emails automatically
- 3: Test your model, repeat steps 1 and 2 until it is good enough
- 4: Launch the relatively easy to mantain model
- Additional: Fully automate the whole process/pipeline





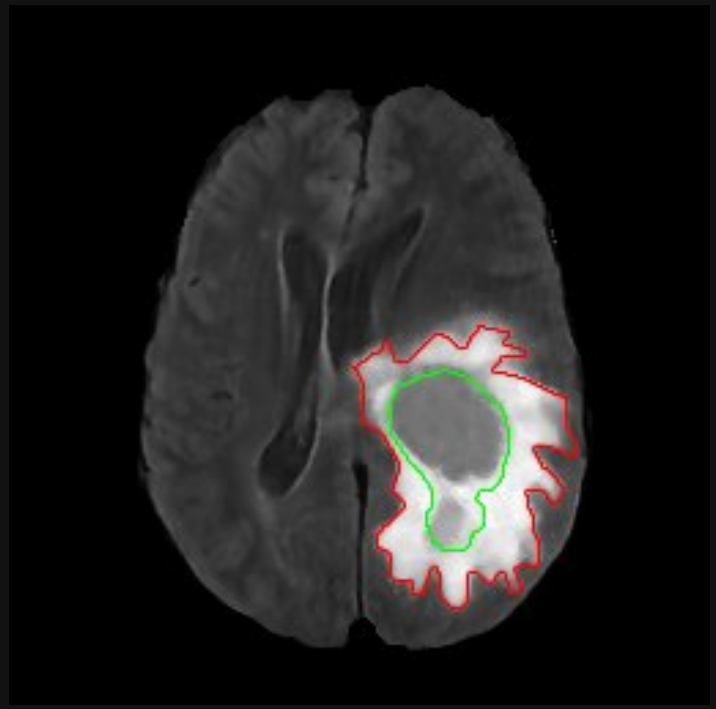
Examples of Applications

- Analyzing images of products on a production line to automatically classify/locate them:
  - Image classification/Object detection
  - CNN - Convolutional Neural Network



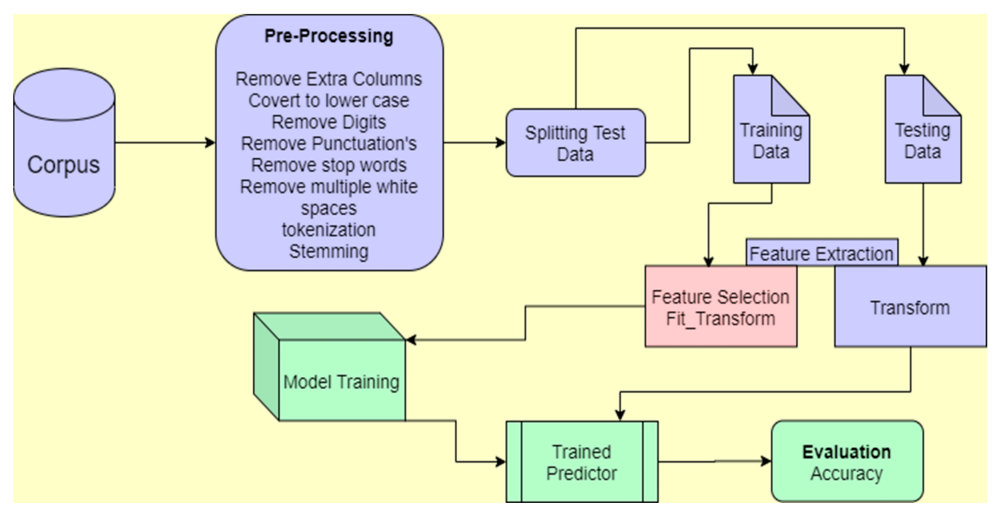
Examples of Applications

- Detecting tumors in brain scans:
  - Segmentation (each pixel is classified as tumor or not tumor)
  - CNN - Convolutional Neural Network or Transformer



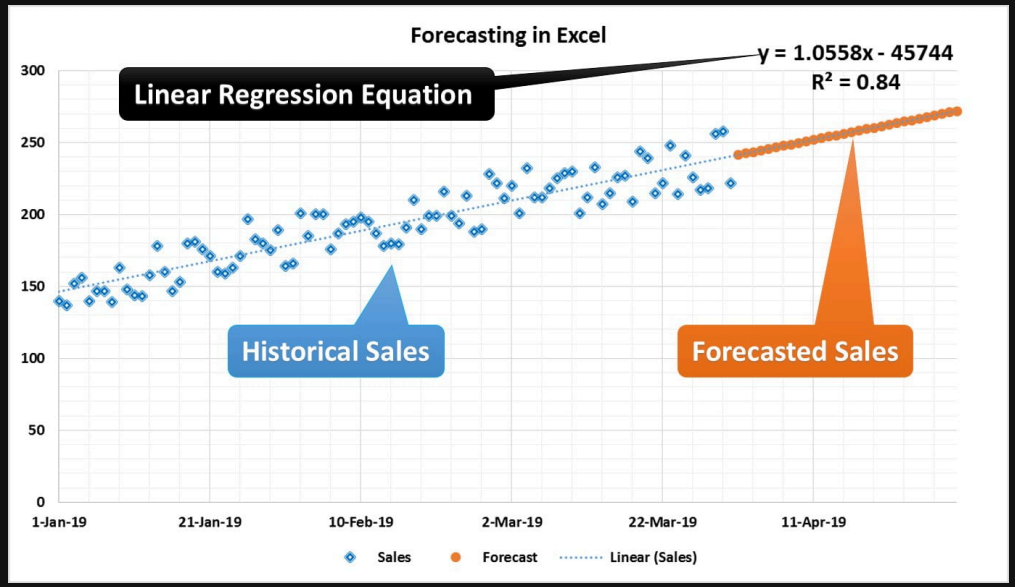
Examples of Applications

- Automatically classifying news articles:
  - Text classification (Natural Language Processing)
  - RNN - Recurrent Neural Network or Transformer



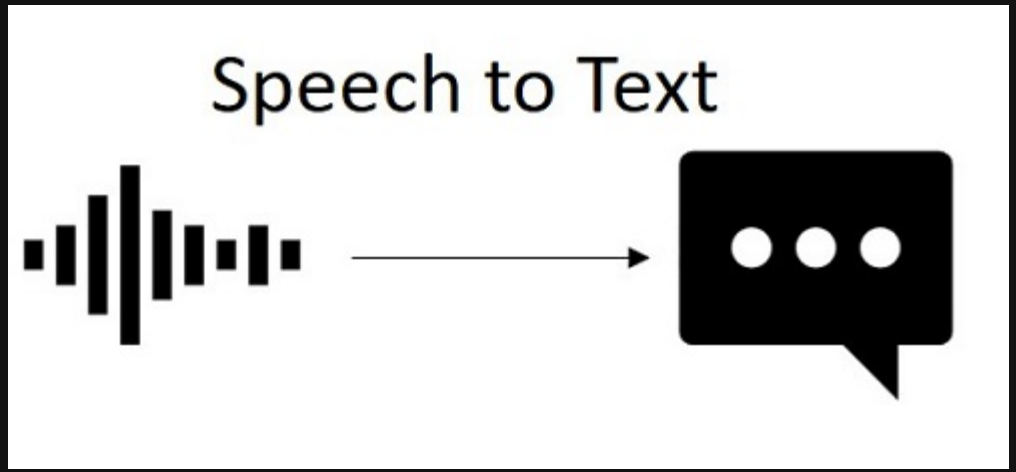
Examples of Applications

- Forecasting your company's revenue next year, based on many performance metrics:
  - Regression (i.e., predicting values)
  - Linear/Polynomial Regression, Neural Network



Examples of Applications

- Making your app react to voice commands:
  - Speech recognition
  - CNN - Convolutional Neural Network, RNN - Recurrent Neural Network, or Transformer



Summarization

- Machine Learning is great for:

  - Problems for which existing solutions require a lot of fine-tuning or long lists of rules
    (a Machine Learning algorithm can often simplify code and perform better)
    - Spam Filter: Long list of rules

  - Complex problems for which there is no good solution at all using a traditional approach
    (a Machine Learning technique can perhaps find a solution)
    - Image recognition: Extremely difficult to write explicit rules for

  - Fluctuating environments - a Machine Learning system can adapt to new data.
    - Self-driving cars: Needs to adapt to new situations

  - Getting insights about complex problems and large amounts of data.
    - Customer segmentation: Diffucult to identify meaningful customer segments

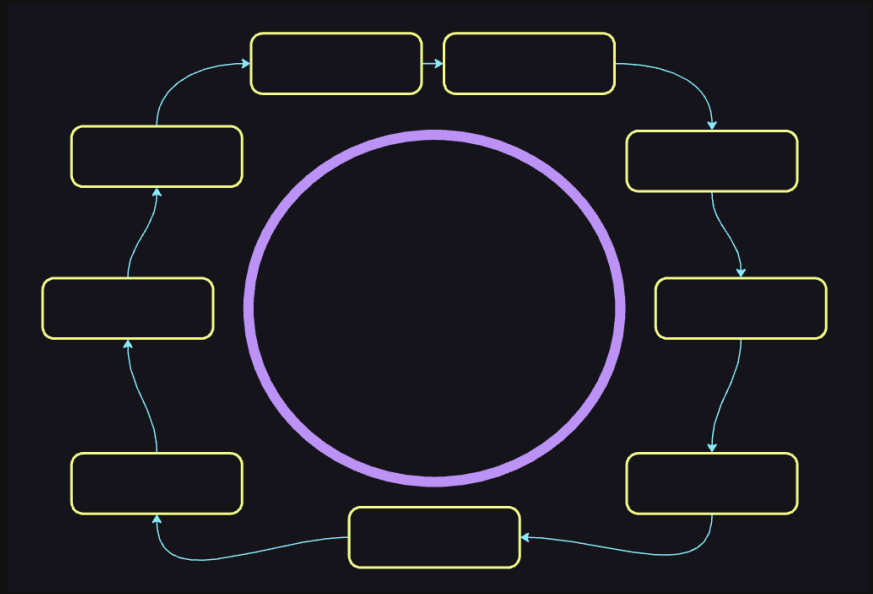
Content

```
Introduction to Machine Learning:
  - What is Machine Learning?
  - Machine Learning Project Checklist
Intro´s:
  - Colab
  - Python
  - Numpy
  - Matplotlib
Supervised Learning:
  - Classification (Binary|Multiclass)
  - Regression
  - Support Vector Machines
  - Decision Trees (Random Forest)
Unsupervised Learning:
  - Dimensionality Reduction
  - Clustering (k-means, DBSCAN)
Performances Measures:
  - Accuracy
  - Precision
  - Recall
  - F1-Score
  - ROC Curve
  - Confusion Matrix
```

Machine Learning Project Checklist

Roadmap (9 steps)

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂️
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙️
- 8) Present 🗣️
- 9) Deploy 🚢



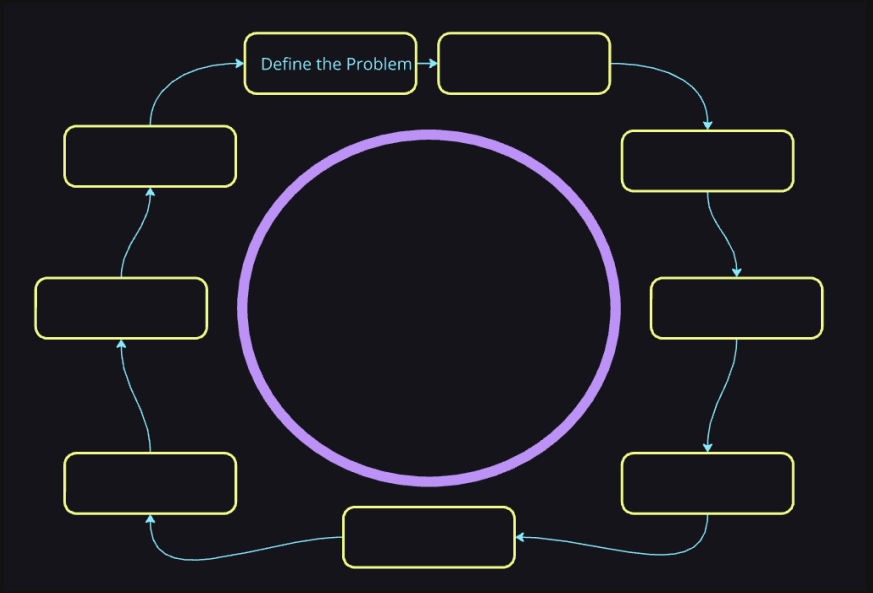
Machine Learning Project Checklist

- 1/9. Define the problem 🎯

Purpose: Understand the business or research problem to solve with ML.

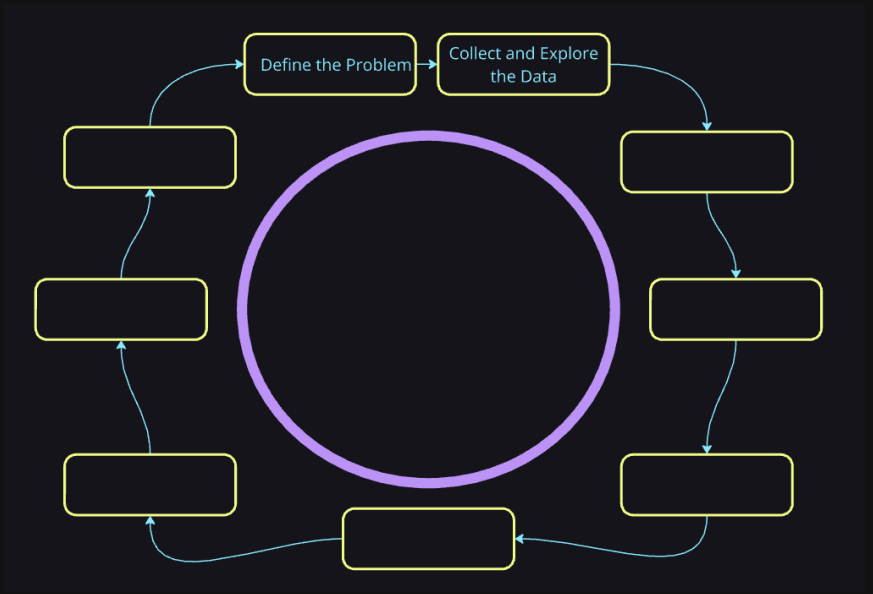
Key Points:
- Identify the problem and the goal (e.g., predict customer churn, classify emails).
- Determine data availability and expected output type.
- Review existing solutions for insights and benchmarks.

Context: Focusing on a specific problem and outcome helps tailor the data collection, modeling approach, and evaluation metrics.



Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



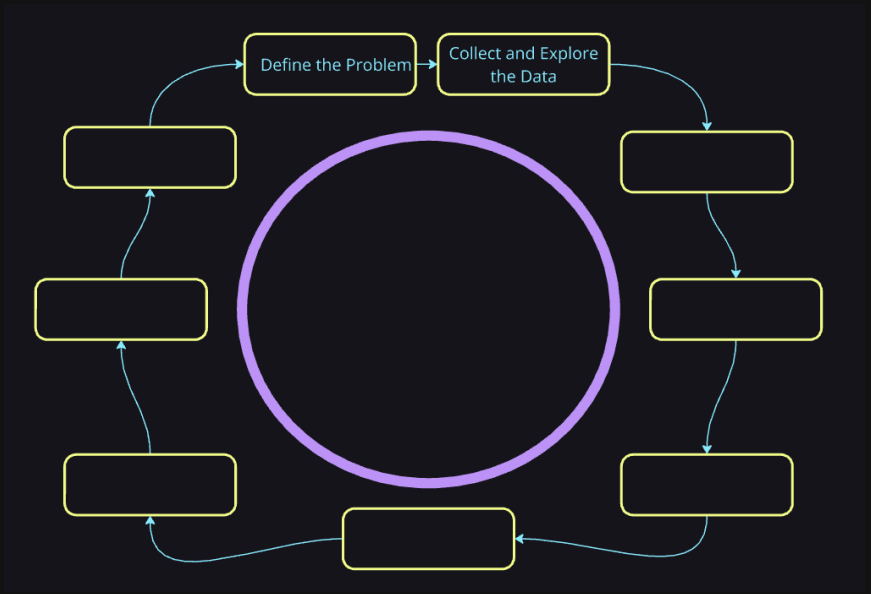
Machine Learning Project Checklist

- 2/9. Collect & explore data 🔎

Purpose: Gather and understand the data necessary for training the model.

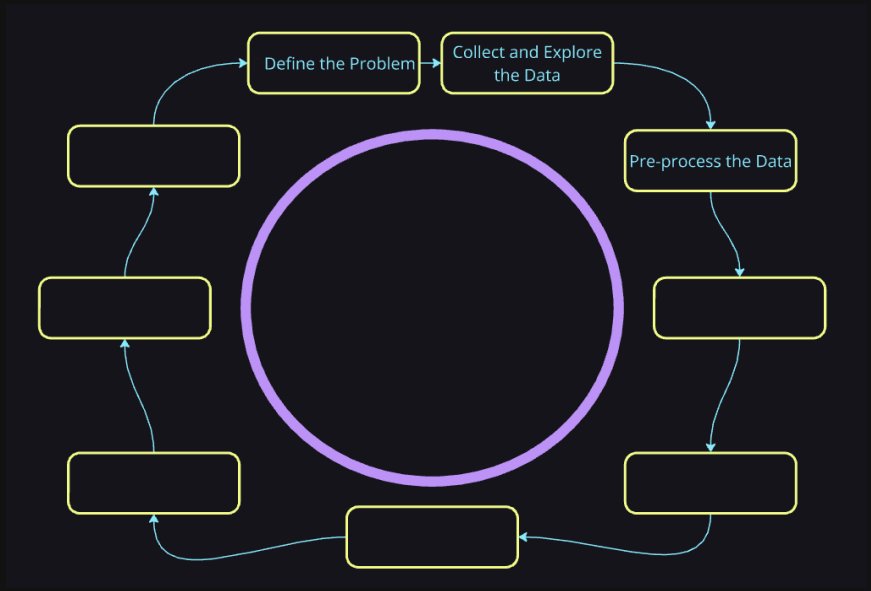
Key Points:
- Sources: Databases, APIs, web scraping.
- Data quality and relevance are crucial for model performance.
- Use statistical tools and visualizations (histograms, scatter plots) to explore data.

Context: Proper data understanding helps identify issues like missing values or outliers and understand variable relationships.



Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



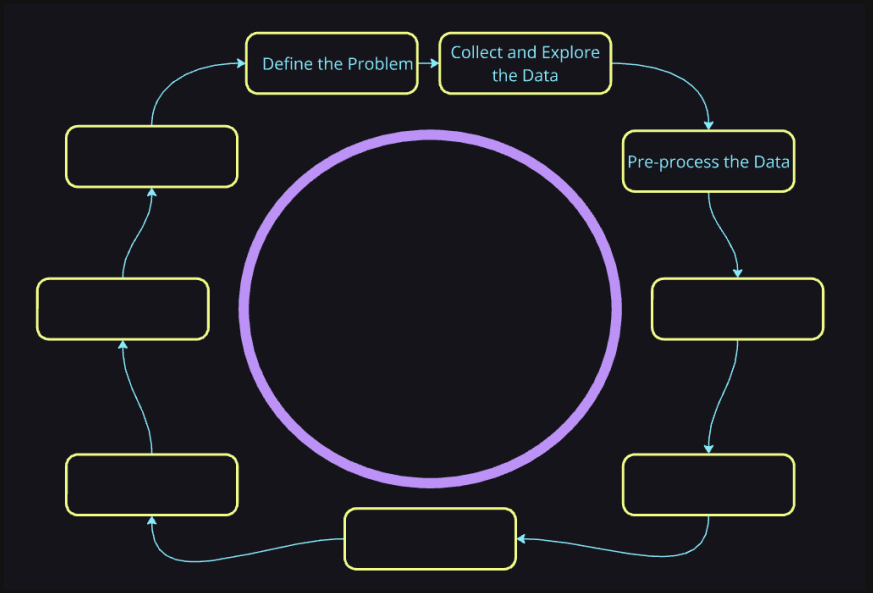
Machine Learning Project Checklist

- 3/9. Pre-process 🧹

Purpose: Prepare the data for machine learning algorithms.

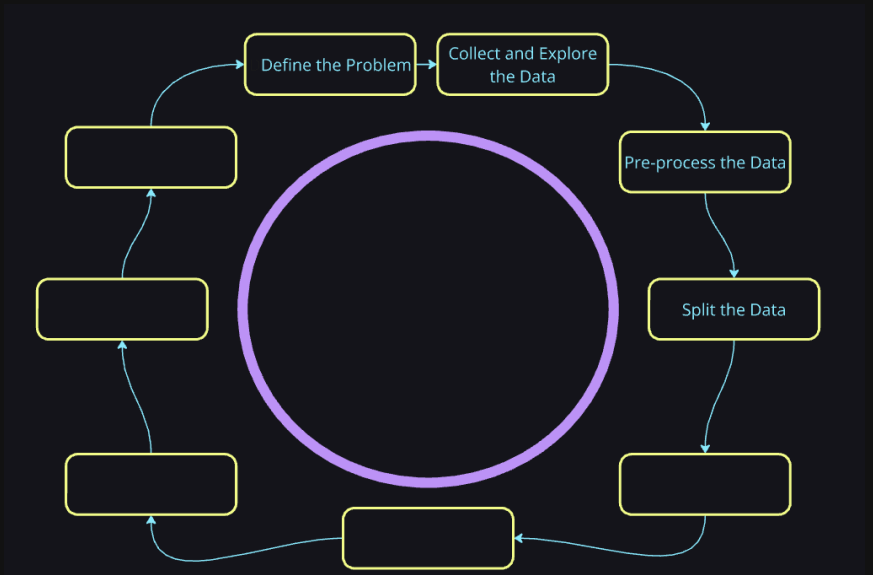
Key Points:
- Address missing values through imputation or exclusion.
- Scale or normalize data to treat all features equally.
- Encode categorical variables (one-hot, label encoding).

Context: These steps are vital to avoid biases and improve algorithm performance.



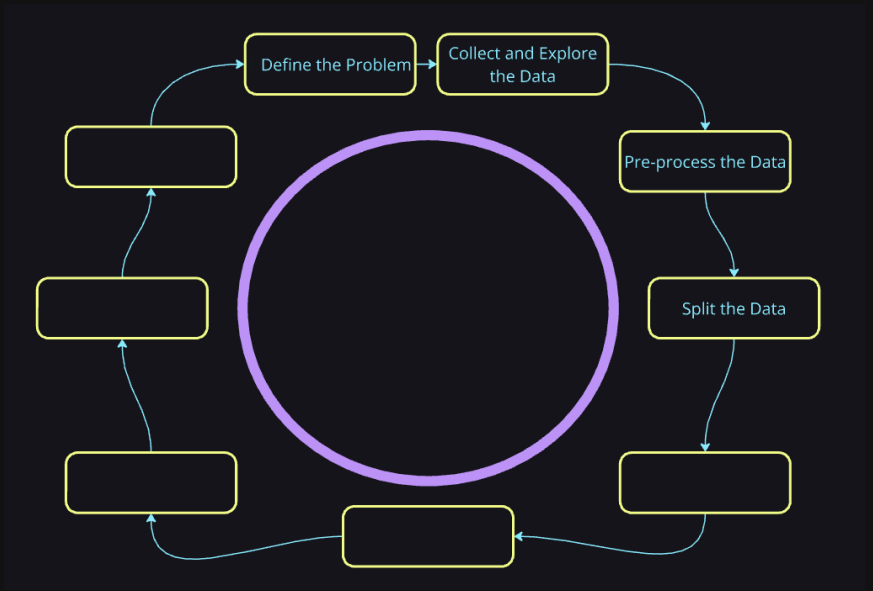
Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



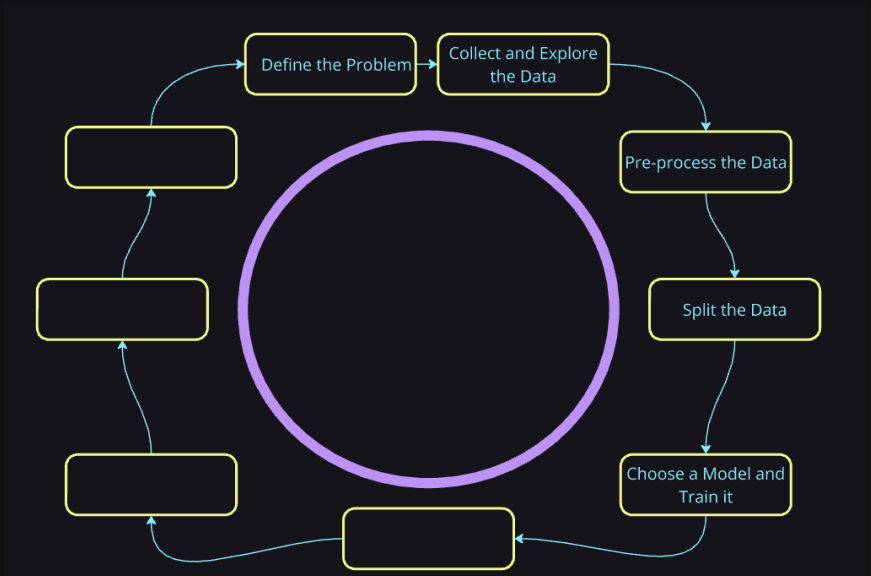
Machine Learning Project Checklist

```
- 4/9. Train/test split ✂

Purpose: Evaluate the model's performance reliably.

Key Points:
    - Training set for model learning.
    - Test set for unbiased evaluation.

Context: Splitting the data helps verify
         the model works well on unseen data,
         ensuring reliability.
```



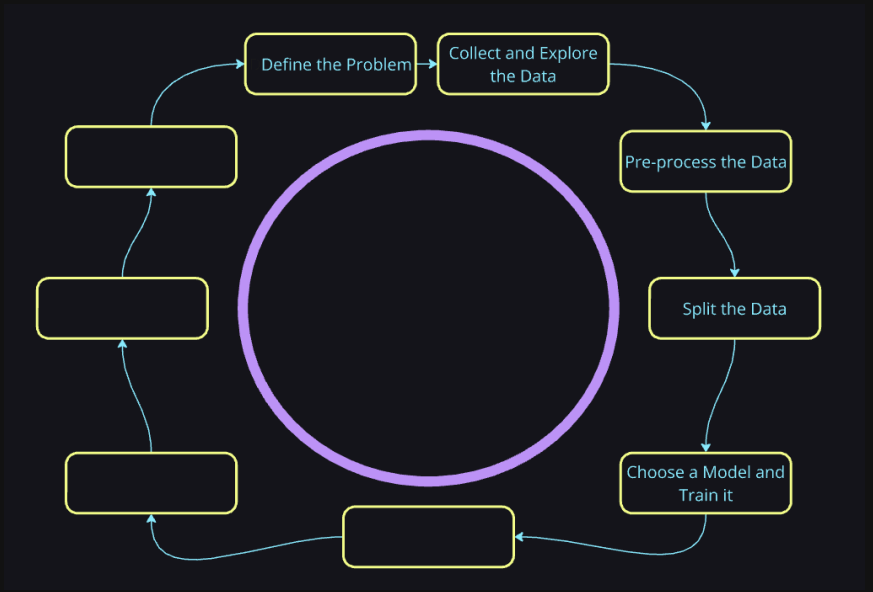
Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



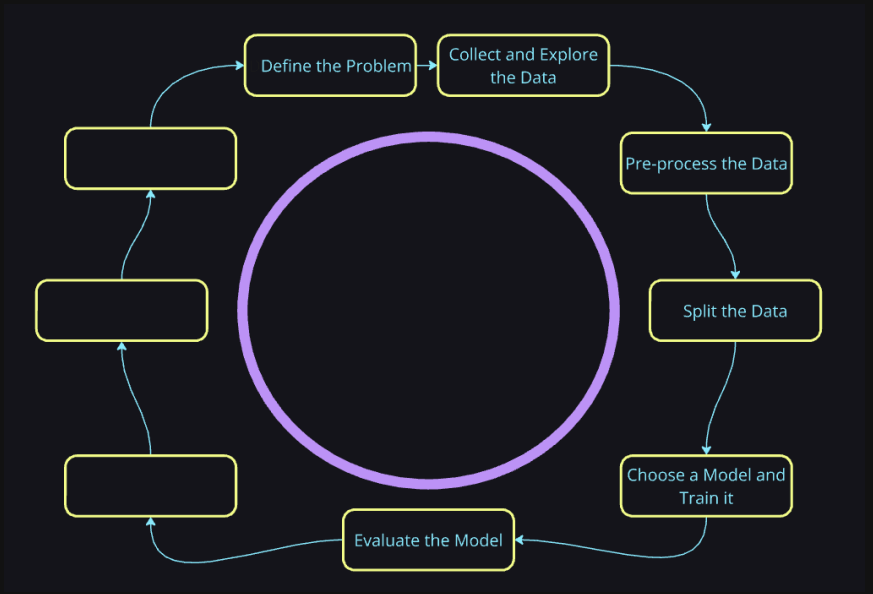
Machine Learning Project Checklist

```
- 5/9. Model & train 🤖

Purpose: Evaluate the model's performance reliably.

Key Points:
    - Model types: Regression, classification,
                   decision trees, neural networks.
    - Train the model on the training data
      to learn from patterns.

Context: Selecting the right model is crucial for
         the model works well on unseen data,
         ensuring reliability.
```



Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



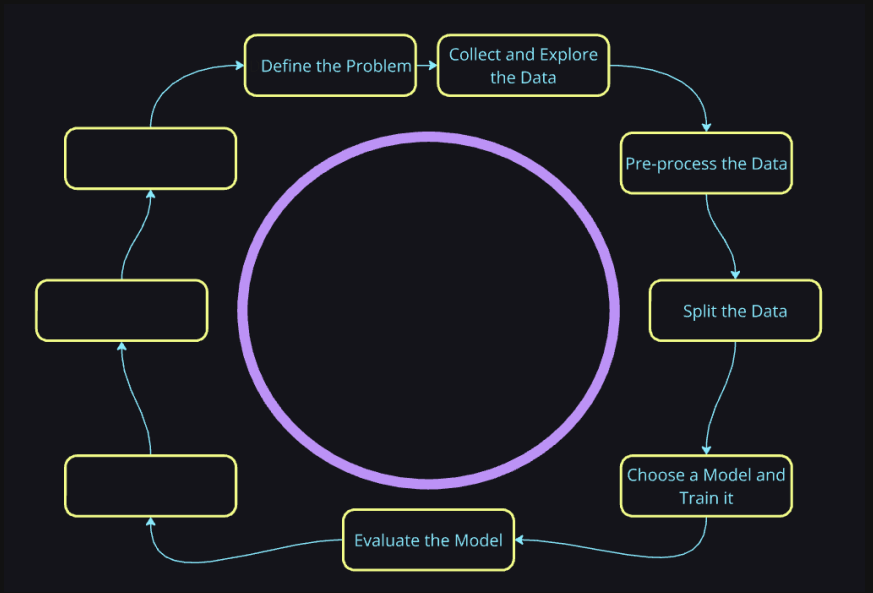
Machine Learning Project Checklist

- 6/9. Evaluate 📊

Purpose: Assess model performance and identify improvement areas.

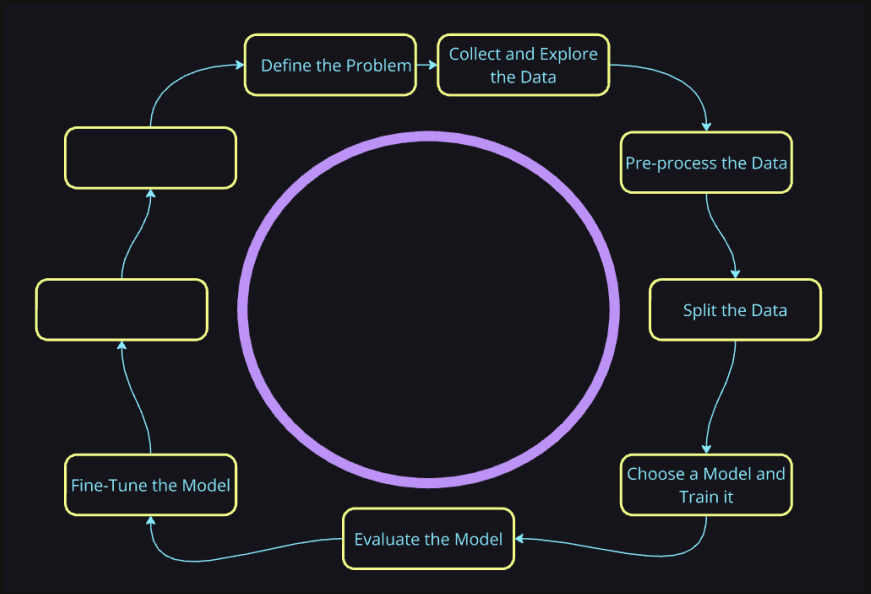
Key Points:
- Use test data for evaluation.
- Metrics: Precision, recall, F1 score, AUC-ROC.

Context: Evaluation confirms the model's effectiveness in practical scenarios and helps fine-tune performance.



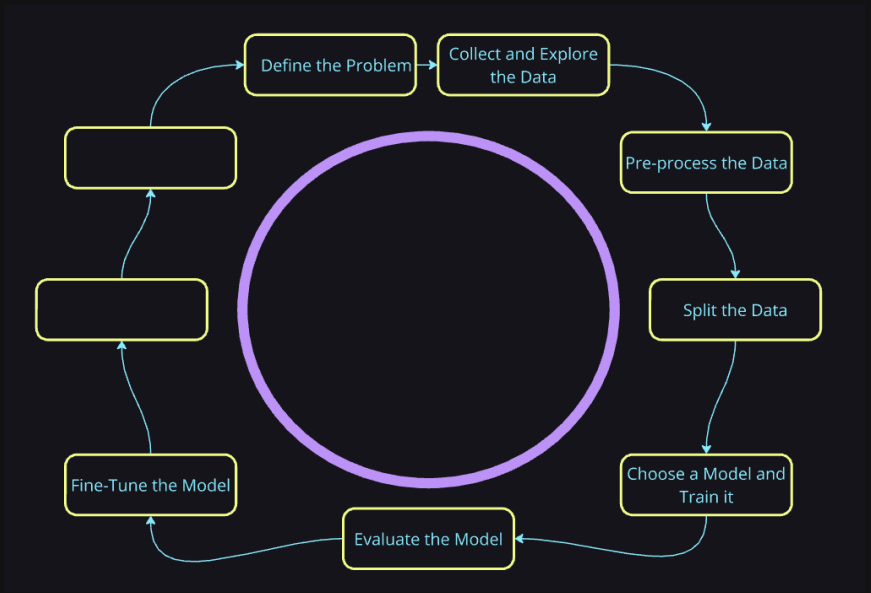
Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



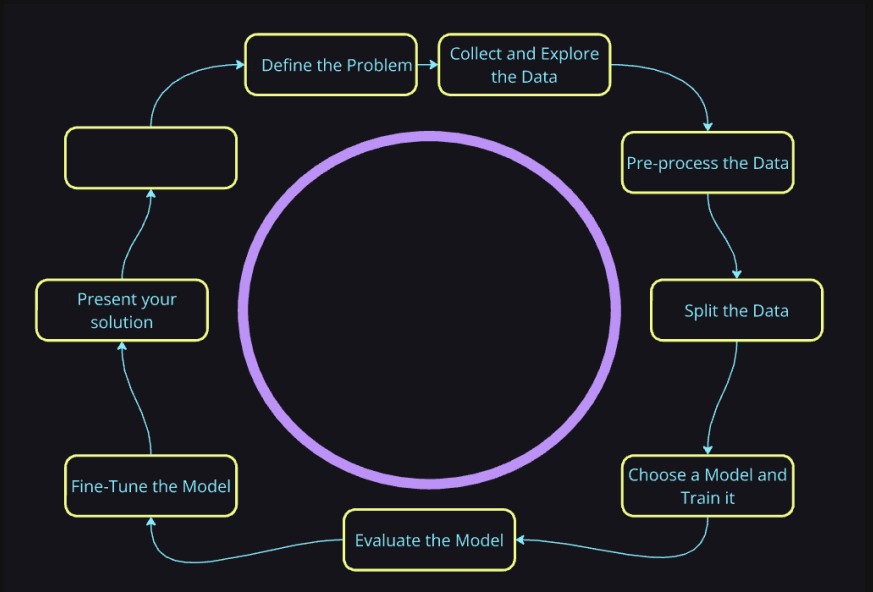
Machine Learning Project Checklist

```
- 7/9. Fine-tune ⚙

Purpose: Optimize the model to achieve the
         best performance.

Key Points:
    - Adjust hyperparameters (learning rate, tree depth).
    - Use techniques like grid search to find
      optimal settings.

Context: Fine-tuning involves tweaking settings
         to refine model behavior and output.
```



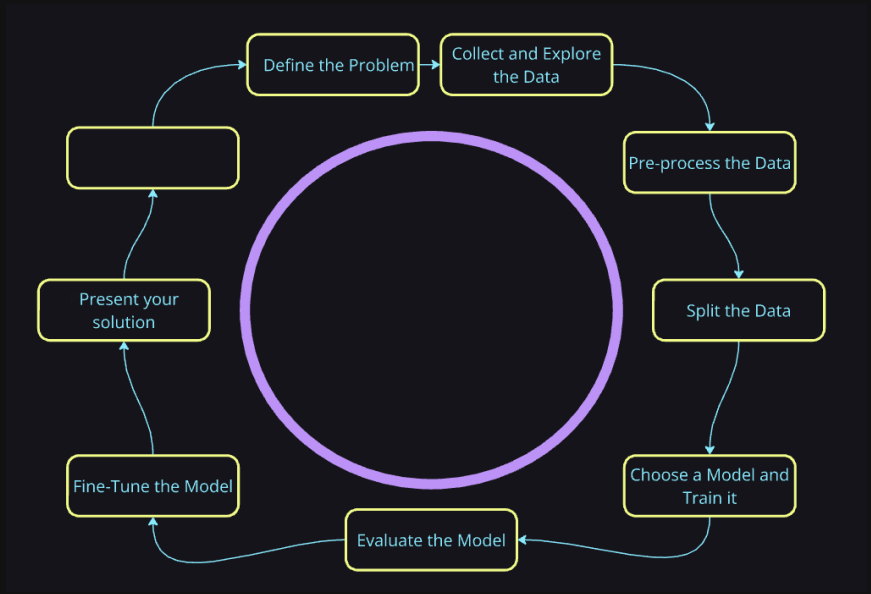
Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



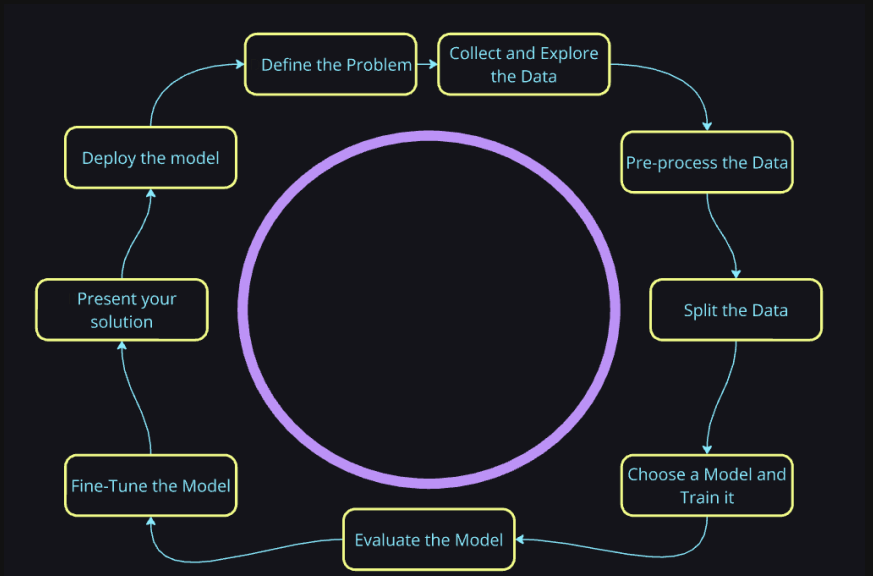
Machine Learning Project Checklist

```
- 8/9. Present 🗣️

Purpose: Communicate the model's findings and implications.

Key Points:
    - Summarize key model insights and performance metrics.
    - Use visualizations to illustrate findings.
    - Recommend actions based on model predictions.

Context: Effective presentation helps stakeholders
         understand and act on model results.
```



Machine Learning Project Checklist

- 1) Define the problem 🎯
- 2) Collect & explore data 🔎
- 3) Pre-process 🧹
- 4) Train/test split ✂
- 5) Model & train 🤖
- 6) Evaluate 📊
- 7) Fine-tune ⚙
- 8) Present 🗣
- 9) Deploy 🚢



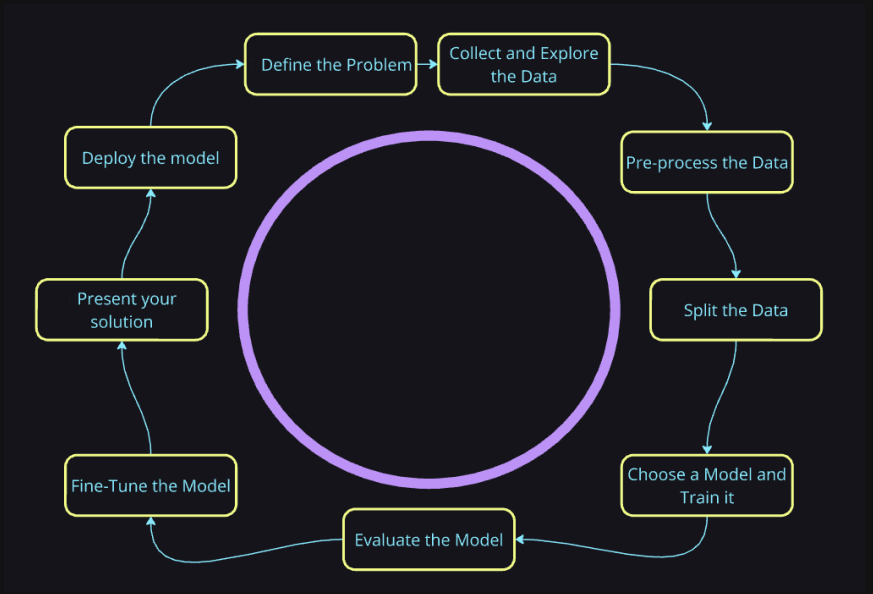
Machine Learning Project Checklist

- 9/9. Deploy 🚢

Purpose: Put the model into production to solve real-world problems.

Key Points:
- Ensure model meets performance and reliability standards.
- Integrate into existing systems or deploy as a new application.

Context: Deployment makes the model available for practical use, impacting decision-making or automation based on its insights.



Machine Learning Project Checklist

```
Overview:
  This checklist provides a structured approach to machine learning projects,
  encompassing all critical phases from problem definition to deployment.

  Key Steps:
  - Define the problem 🎯: Identify what you are trying to solve.
  - Collect & explore data 🔎: Gather necessary data and explore its characteristics.
  - Pre-process 🧹 : Clean and prepare data for modeling.
  - Train/test split ✂️: Divide data into training and test sets.
  - Model & train 🤖: Select an appropriate model and train it on your data.
  - Evaluate 📊: Assess the model's performance on the test set.
  - Fine-tune ⚙️: Optimize the model's settings and parameters.
  - Present 🗣️: Summarize findings and suggest actions.
  - Deploy 🚢: Implement the model in a real-world environment.
```

Content

```
Introduction to Machine Learning:
  - What is Machine Learning?
  - Machine Learning Project Checklist
Intro´s:
  - Colab
  - Python
  - Numpy
  - Matplotlib
Supervised Learning:
  - Classification (Binary|Multiclass)
  - Regression
  - Support Vector Machines
  - Decision Trees (Random Forest)
Unsupervised Learning:
  - Dimensionality Reduction
  - Clustering (k-means, DBSCAN)
Performances Measures:
  - Accuracy
  - Precision
  - Recall
  - F1-Score
  - ROC Curve
  - Confusion Matrix
```