Génie logiciel pour la conception d'un Système d'Information

CSC4521

Voie d'Approfondissement
Intégration et Déploiement de Systèmes d'Information
(VAP DSI)

AI - Machine learning

http://jpaulgibson.synology.me/~jpaulgibson/TSP/Teaching/CSC4521/

.../CSC4521/CSC4521-AI-ML.pdf

In your Wavestone projects you have all identified ***Algorithmic AI*** components in your systems

You need to be precise what you mean:

1) Traditional deterministic procedural algorithm
2) Expert systems : knowledge base and inference engine - declarative programming
3) Machine Learning - no explicit program

There are advantages and negatives of each of these

Do not assume that ML is the best choice for your needs

You must analyse what function you want your AI to calculate, … as well as many other factors

With ML there are many alternative approaches

ETHICAL ASPECTS TO AI, GEN-AI USE

A reminder -

- The environment - energy, water, pollution
- Intellectual Property - theft, plagiarism
- Human labour exploitation/abuse
- ‘Hallucinations’
- Bias and discrimination
- Data privacy/security
- Fake news - disinformation, fraud, & scams
- Fake media - abuse and illegal images/videos/text
- Cognitive and Technical Debt
- Psychological impacts - anthropomorphism, nudging and manipulation
- Destruction/pollution of the WWW
- Economic risks - size of investments, undermining creative economies
- Tech. Resource risks - other critical technological uses/advances are being slowed

future internet

MDPI

Review

Machine Learning: Models, Challenges, and Research Directions

Tala Talaei Khoei * and Naima Kaabouch

School of Computer Science and Electrical Engineering, University of North Dakota, Grand Forks, ND 58202, USA; naima.kaabouch@und.edu
* Correspondence: tala.talaeikhoei@und.edu

Abstract: Machine learning techniques have emerged as a transformative force, revolutionizing various application domains, particularly cybersecurity. The development of optimal machine learning applications requires the integration of multiple processes, such as data pre-processing, model selection, and parameter optimization. While existing surveys have shed light on these techniques, they have mainly focused on specific application domains. A notable gap that exists in current studies is the lack of a comprehensive overview of machine learning architecture and its essential phases in the cybersecurity field. To address this gap, this survey provides a holistic review of current studies in machine learning, covering techniques applicable to any domain. Models are classified into four categories: supervised, semi-supervised, uns… learning. Each of these categories and their models are described. In… the current progress related to data pre-processing and hyperparameter tuning techniques. Moreover, this survey identifies and reviews the research gaps and key challenges that the cybersecurity field faces. By analyzing these gaps, we propose some promising research directions for the future. Ultimately, this survey aims to serve as a valuable resource for researchers interested in learning about machine learning, providing them with insights to foster innovation and progress across diverse application domains.

Keywords: artificial intelligence; data pre-processing; machine learning; supervised learning; semi-supervised learning; optimization techniques; reinforcement learning; unsupervised learning

Embedded PDF

check for updates

Citation: Talaei Khoei, T.; Kaabouch, N. Machine Learning: Models, Challenges, and Research Directions. *Future Internet* **2023**, *15*, 332. https://doi.org/10.3390/fi15100332

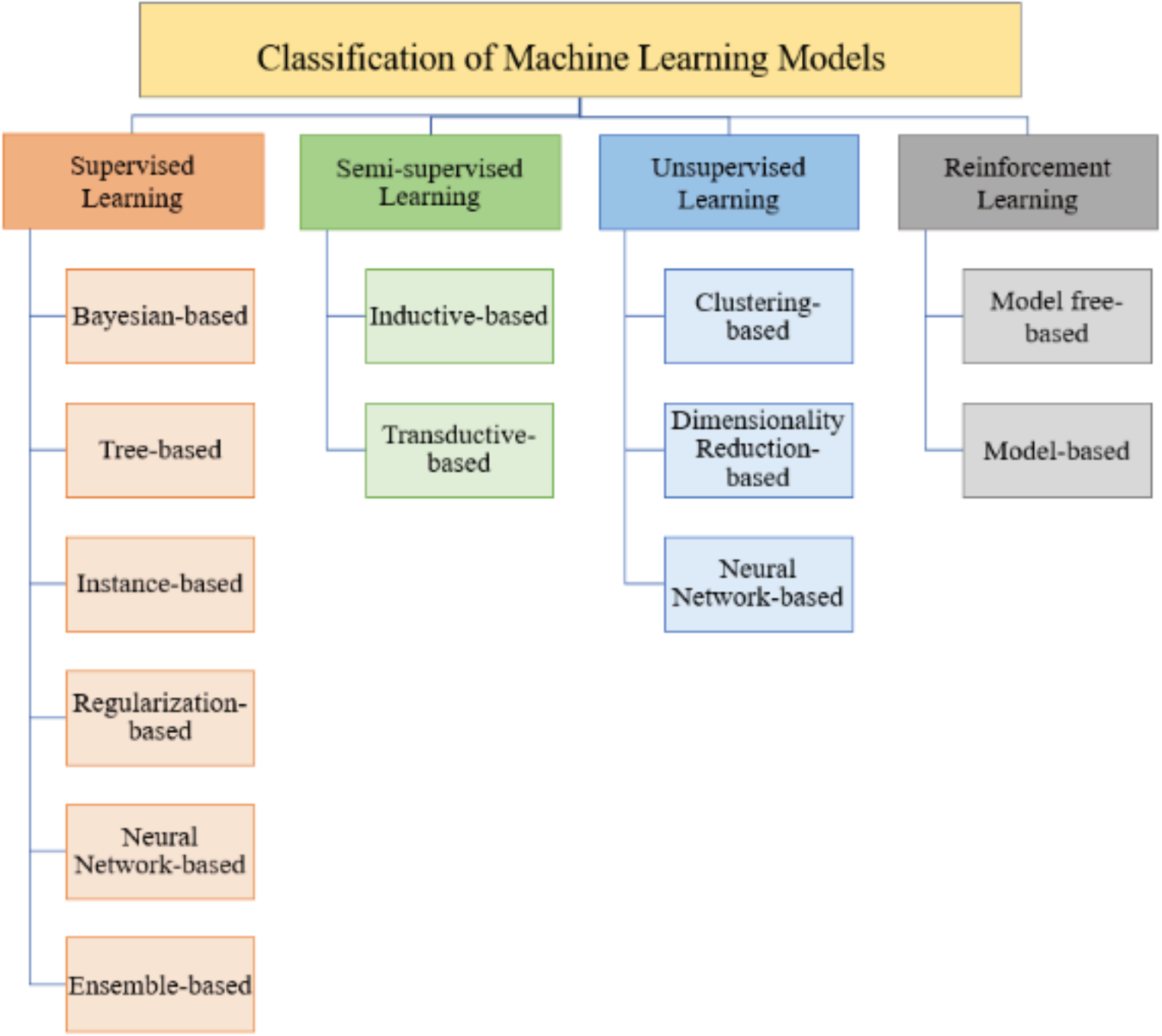
Academic Editor: Massimo Cafaro

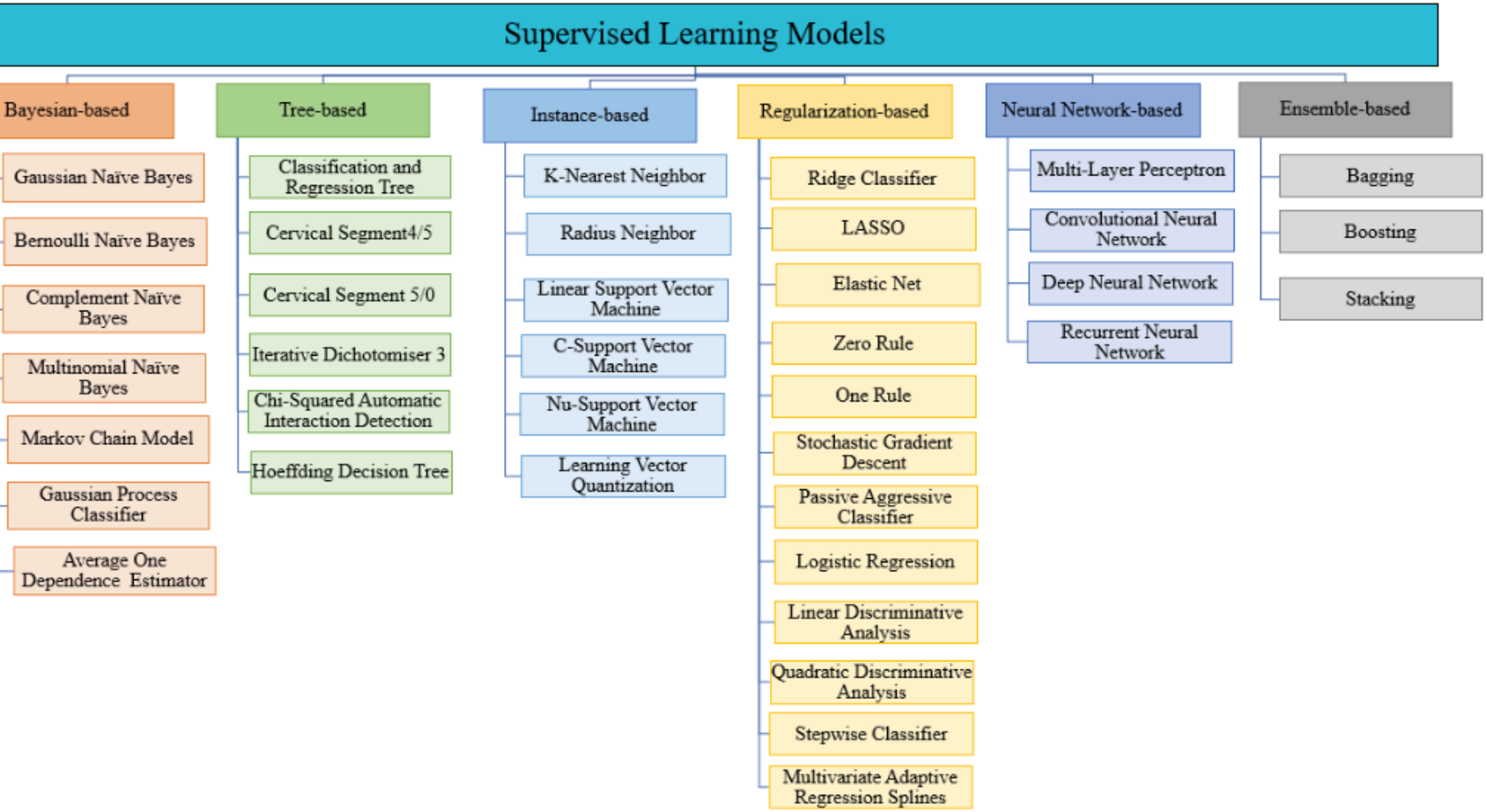
Received: 8 September 2023
Revised: 25 September 2023
Accepted: 3 October 2023
Published: 9 October 2023

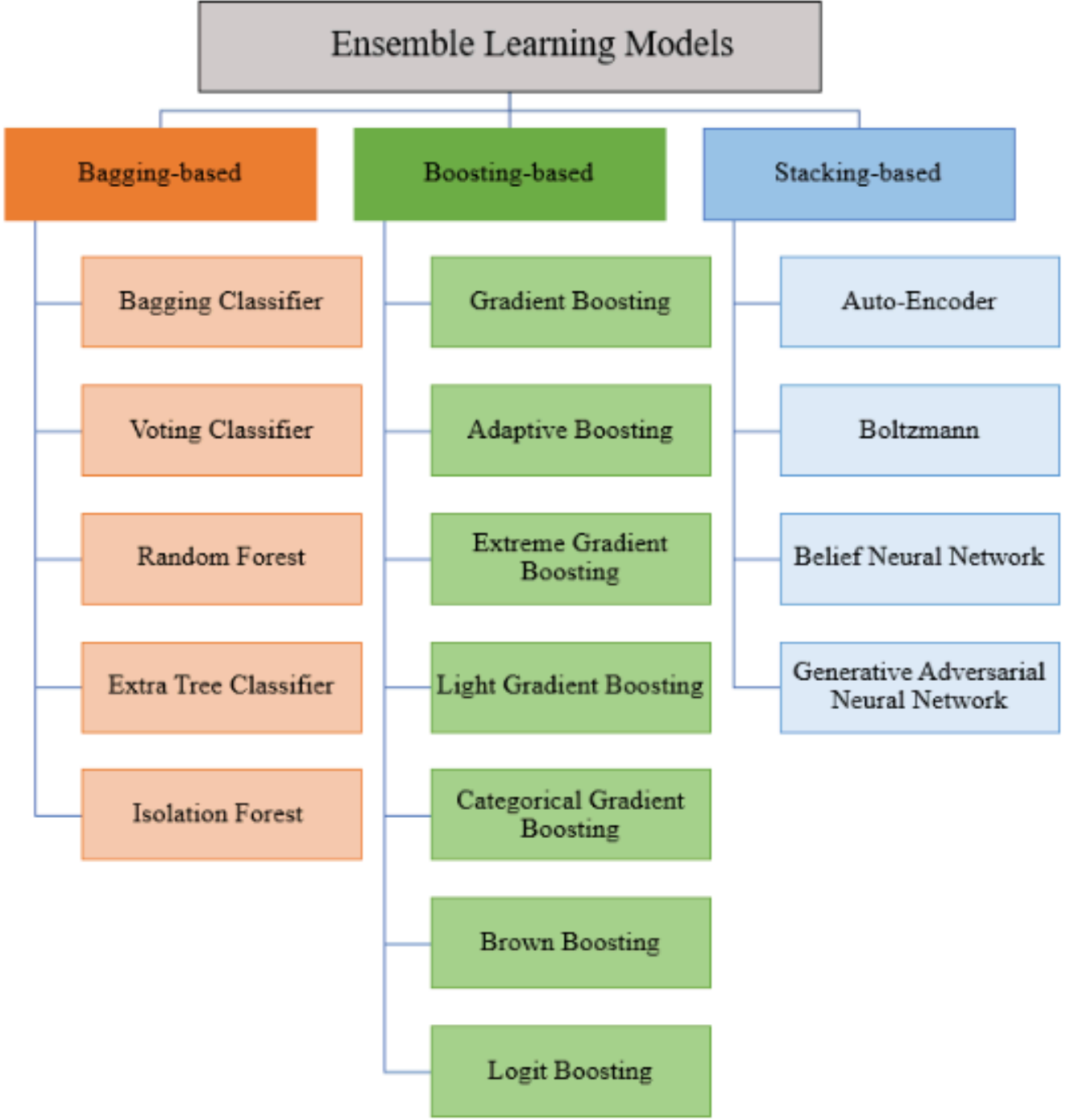
1. Introduction

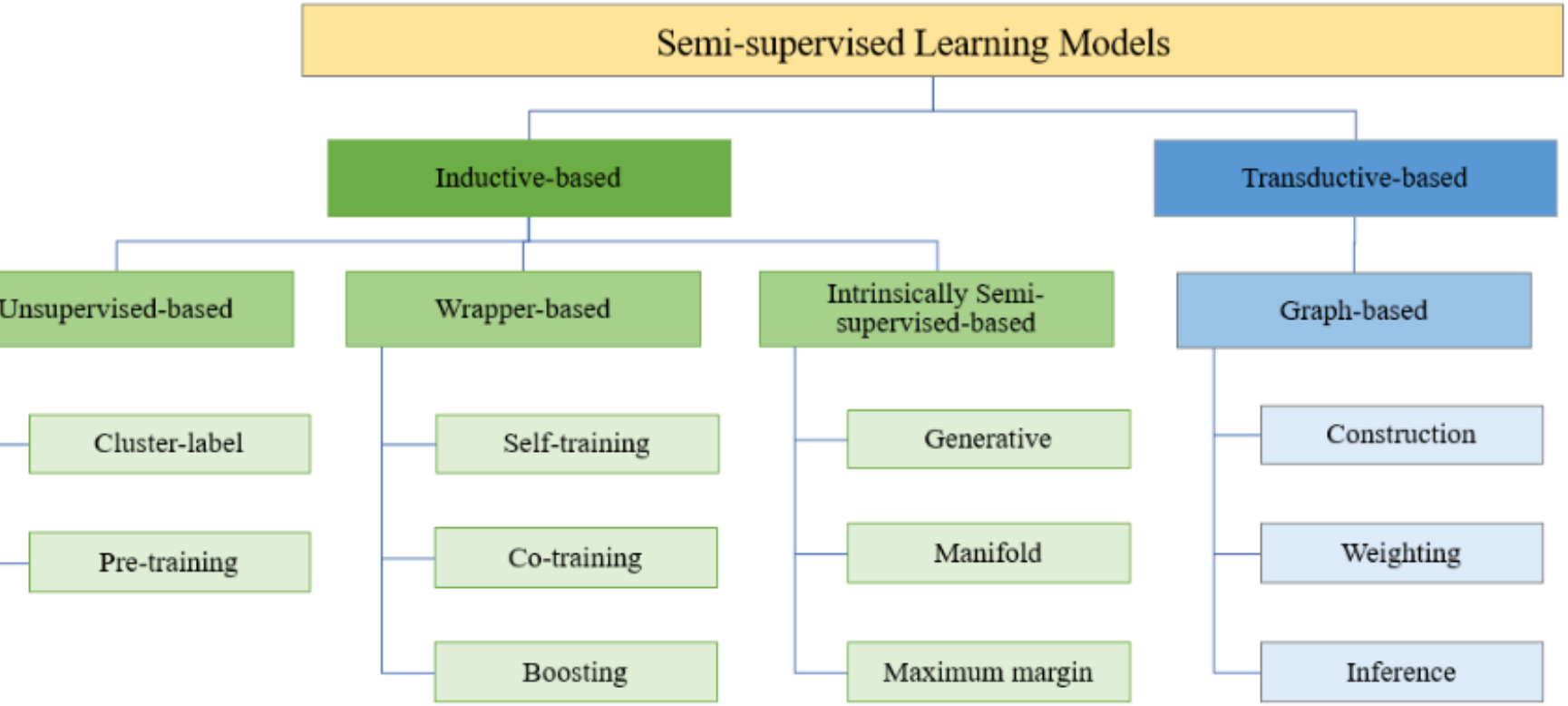
Machine learning (ML) applications have recently found extensive use across various domains, including cyber-security [1]. These cyber-security applications have attained widespread adoption in research and commercial systems, establishing themselves as indispensable components [2,3]. Building a compelling and influential secure ML application is a complex and time-consuming undertaking. It requires high-quality data and the selection of an appropriate model, which can achieve optimal architecture through hyperparameter tuning [4]. Numerous factors have to be taken into consideration when building a successful machine learning model. To develop a thriving security ML application, the initial step involves collecting raw signals and generating a dataset. However, this process often encounters challenges, such as inconsistencies, imputations, and redundancies, which can lead to inaccurate results [5–7]. Therefore, employing proper data pre-processing techniques becomes imperative in the development of any successful machine learning application [8,9]. The second step is to choose a machine learning model. These ML models can be classified into four categories: supervised, semi-supervised, unsupervised, and reinforcement learning. Supervised models use a labeled dataset for the training process to achieve the desired outputs [3]. In contrast, unsupervised models are trained using unlabeled datasets without any supervision. Semi-supervised models are combinations of supervised and unsupervised models, which train the model using a small number

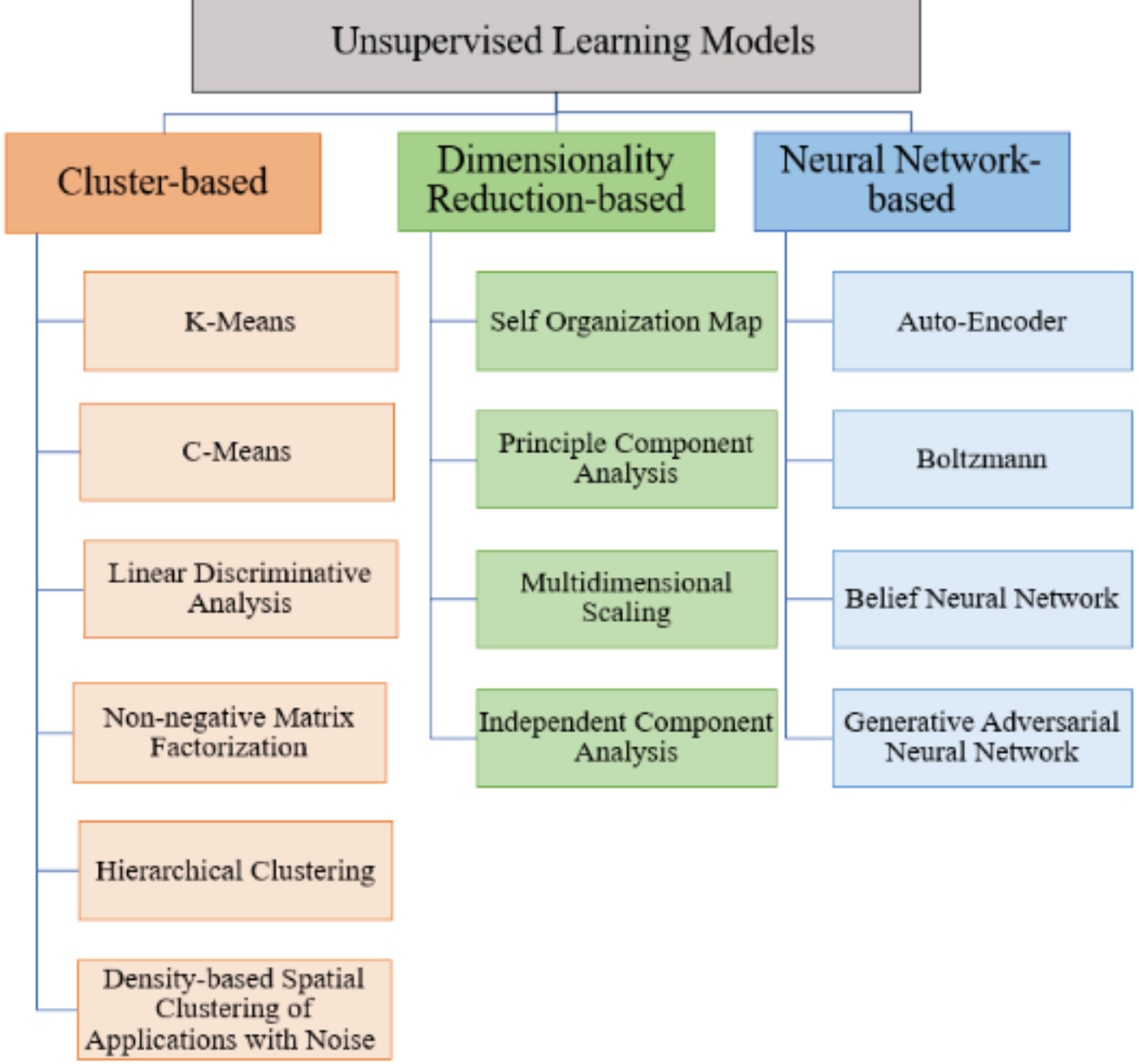
Talaei Khoei, Tala, and Naima Kaabouch. "Machine learning: Models, challenges, and research directions." *Future Internet* 15.10 (2023): 332.

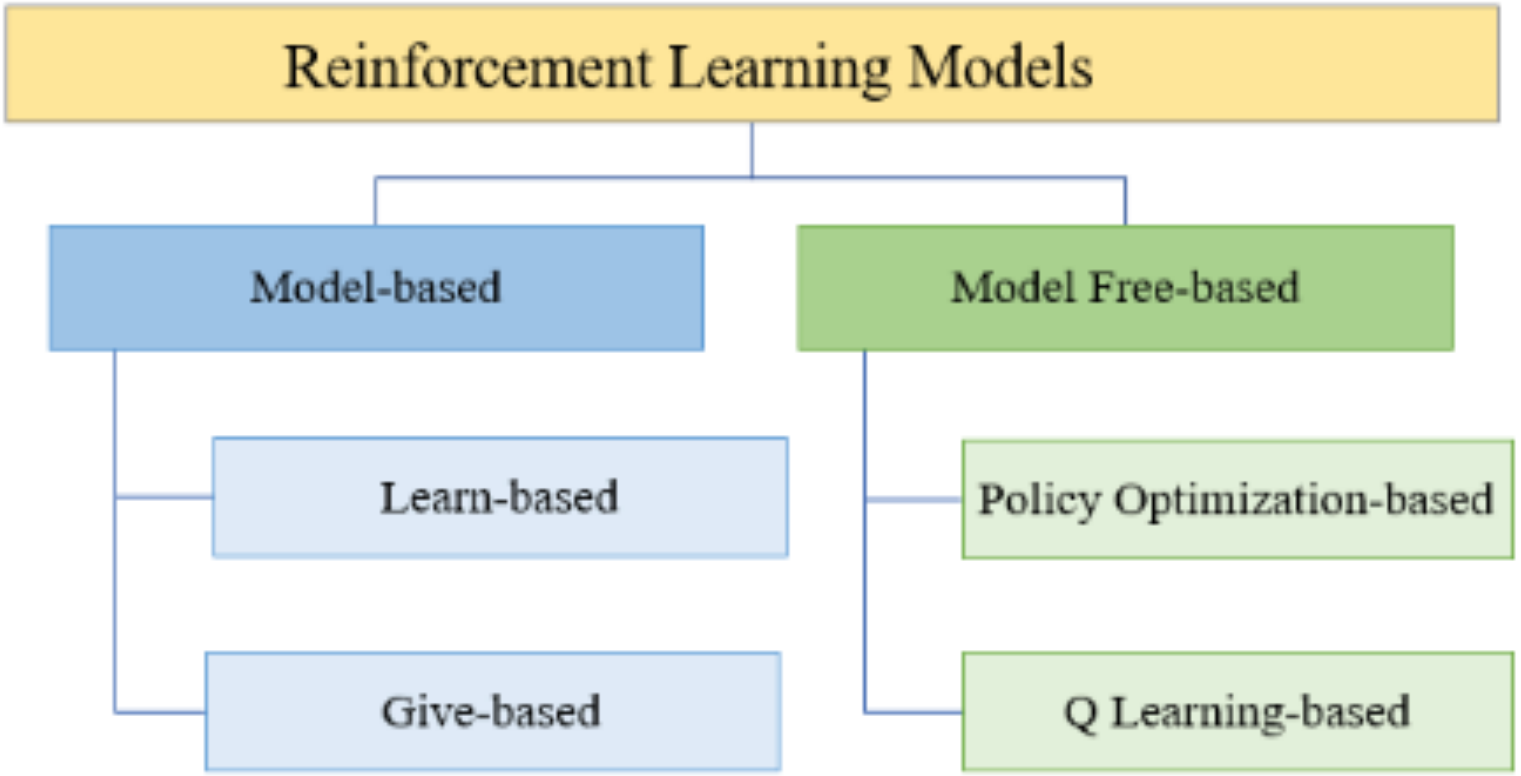

Classification of Machine Learning Models
Supervised Learning
Semi-supervised Learning
Unsupervised Learning
Reinforcement Learning
Bayesian-based
Tree-based
Instance-based
Regularization-based
Neural Network-based
Ensemble-based
Inductive-based
Transductive-based
Clustering-based
Dimensionality Reduction-based
Neural Network-based
Model free-based
Model-based


Supervised Learning Models
Bayesian-based
Gaussian Naïve Bayes
Bernoulli Naïve Bayes
Complement Naïve Bayes
Multinomial Naïve Bayes
Markov Chain Model
Gaussian Process Classifier
Average One Dependence Estimator
Tree-based
Classification and Regression Tree
Cervical Segment4/5
Cervical Segment 5/0
Iterative Dichotomiser 3
Chi-Squared Automatic Interaction Detection
Hoeffding Decision Tree
Instance-based
K-Nearest Neighbor
Radius Neighbor
Linear Support Vector Machine
C-Support Vector Machine
Nu-Support Vector Machine
Learning Vector Quantization
Regularization-based
Ridge Classifier
LASSO
Elastic Net
Zero Rule
One Rule
Stochastic Gradient Descent
Passive Aggressive Classifier
Logistic Regression
Linear Discriminative Analysis
Quadratic Discriminative Analysis
Stepwise Classifier
Multivariate Adaptive Regression Splines
Neural Network-based
Multi-Layer Perceptron
Convolutional Neural Network
Deep Neural Network
Recurrent Neural Network
Ensemble-based
Bagging
Boosting
Stacking


Ensemble Learning Models
Bagging-based
Boosting-based
Stacking-based
Bagging Classifier
Voting Classifier
Random Forest
Extra Tree Classifier
Isolation Forest
Gradient Boosting
Adaptive Boosting
Extreme Gradient Boosting
Light Gradient Boosting
Categorical Gradient Boosting
Brown Boosting
Logit Boosting
Auto-Encoder
Boltzmann
Belief Neural Network
Generative Adversarial Neural Network


Semi-supervised Learning Models
Inductive-based
Transductive-based
Unsupervised-based
Wrapper-based
Intrinsically Semi-supervised-based
Graph-based
Cluster-label
Pre-training
Self-training
Co-training
Boosting
Generative
Manifold
Maximum margin
Construction
Weighting
Inference


Unsupervised Learning Models
Cluster-based
Dimensionality Reduction-based
Neural Network-based
K-Means
C-Means
Linear Discriminative Analysis
Non-negative Matrix Factorization
Hierarchical Clustering
Density-based Spatial Clustering of Applications with Noise
Self Organization Map
Principle Component Analysis
Multidimensional Scaling
Independent Component Analysis
Auto-Encoder
Boltzmann
Belief Neural Network
Generative Adversarial Neural Network


Reinforcement Learning Models
Model-based
Model Free-based
Learn-based
Give-based
Policy Optimization-based
Q Learning-based

UC Berkeley School of Information

Cybersecurity Data Science Admissions Experience

Home

What Is Machine Learning (ML)? Definition and Examples

March 17, 2026

https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/

Key Takeaways

- Machine learning is a subset of AI concerned with training models to allow computers to mimic human thought and decision making without explicit programming.
- The most common types of ML are supervised learning (learning via labeled data), unsupervised learning (learning via unlabeled data), and reinforcement learning (learning via a reward and punishment response).
- While ML drives powerful benefits like automation, advanced pattern recognition, and personalization, it also has risks, including algorithmic bias and data privacy concerns.

Three Components of a Machine Learning Model

There are a variety of different ML model architectures and types. However, the typical supervised machine learning algorithm consists of roughly three components:

1. A Decision Process

A decision process is a recipe of calculations or other steps that takes in the data and “guesses” what kind of pattern your algorithm is looking to find.

2. An Error Function

An error function is a method of measuring how good the guess was by comparing it to known examples (when they are available). Did the decision process get it right? If not, how do you quantify “how bad” the miss was?

3. An Updating or Optimization Process

An updating or optimization process is a method in which the algorithm looks at the miss and then updates how the decision process comes to the final decision, so next time the miss won’t be as great.

Benefits and Risks of Machine Learning

Machine learning is transformational, but should be applied with a balanced perspective. While the benefits in efficiency and insight are powerful, the technology is not neutral. When left unchecked, ML can present significant technical and ethical challenges.

Benefits of Machine Learning	Risks of Machine Learning
ML automates time-consuming, repetitive tasks	Algorithmic bias can reinforce stereotypes and harmful outcomes
Models can identify patterns invisible to humans	'Black box' nature of models makes transparency difficult
Continuous iteration improves accuracy and efficiency	Models that require access to sensitive data create security vulnerabilities
Scales and handles large datasets easily	Computational costs can be prohibitive
Enables innovation and advanced solution	Poor data quality can lead to inaccurate predictions

Benefits and Risks of Machine Learning

Machine learning is transformational, but should be applied with a balanced perspective. While the benefits in efficiency and insight are powerful, the technology is not neutral. When left unchecked, ML can present significant technical and ethical challenges.

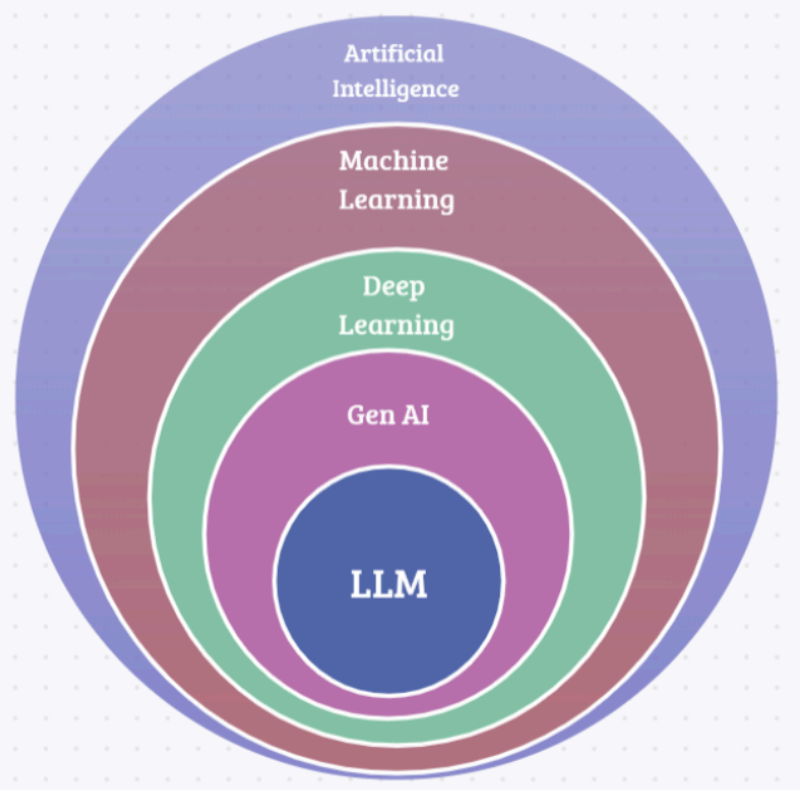
Benefits of Machine Learning	Risks of Machine Learning
ML automates time-consuming, repetitive tasks	Algorithmic bias can reinforce stereotypes and harmful outcomes
Models can identify patterns invisible to humans	'Black box' nature of models makes transparency difficult
Continuous iteration improves accuracy and efficiency	Models that require access to sensitive data create security vulnerabilities
Scales and handles large datasets easily	Computational costs can be prohibitive
Enables innovation and advanced solution	Poor data quality can lead to inaccurate predictions

Generative AI

A generative AI model is a form of machine learning that does more than process data, it creates it. By studying huge sets of examples, these models learn patterns in language, visuals, sound, and more. Then, they use that knowledge to produce new content that mirrors the style and structure of what they have learned.

Generative AI - a simplistic view

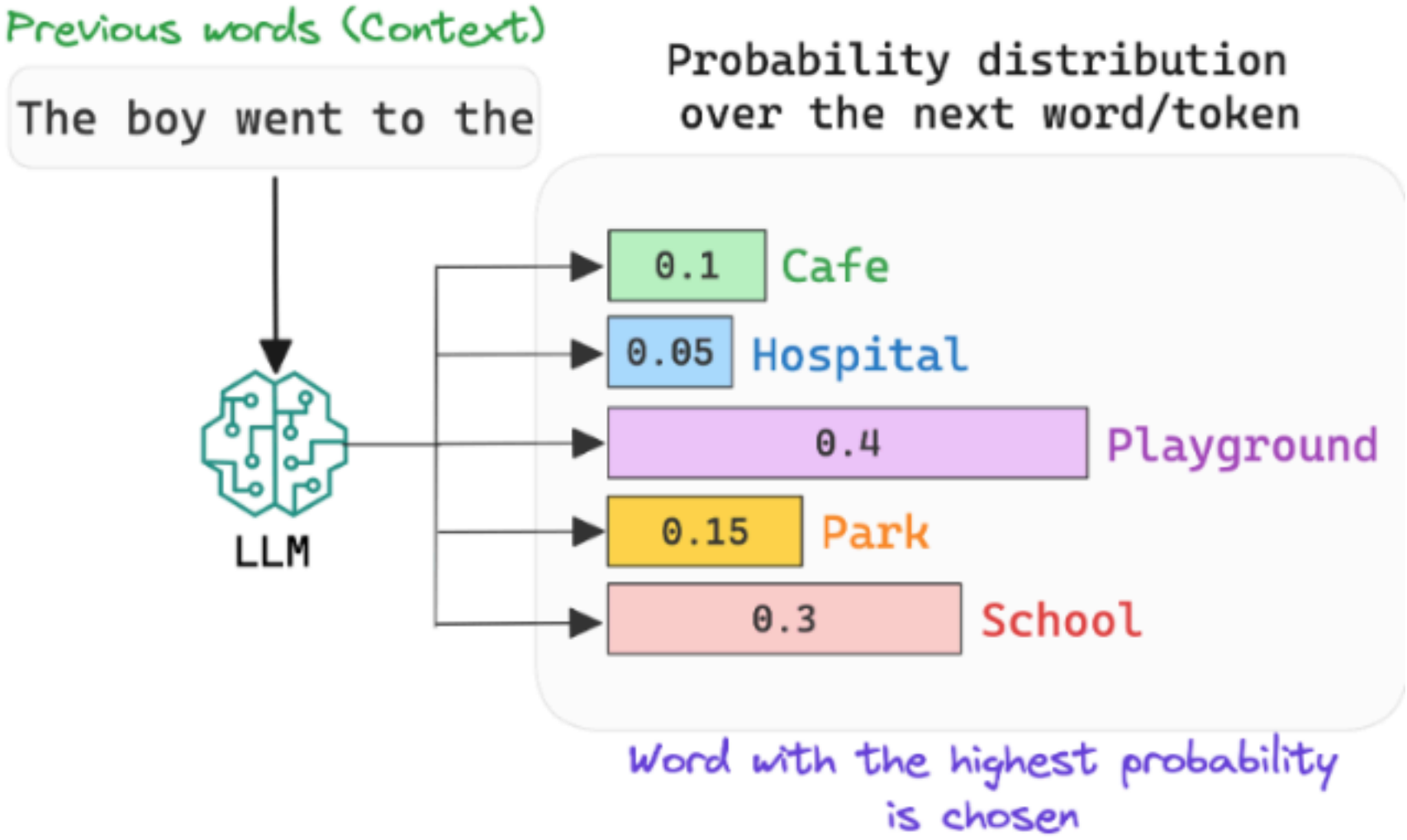
A generative AI model is a form of machine learning that does more than process data, it creates it. By studying huge sets of examples, these models learn patterns in language, visuals, sound, and more. Then, they use that knowledge to produce new content that mirrors the style and structure of what they have learned.



Some Types of Generative AI

1. Generative Adversarial Networks (GANs)
2. Variational Autoencoders (VAEs)
3. Autoregressive Models
4. Recurrent Neural Networks (RNNs)
5. Transformer-based Models
6. Reinforcement Learning for Generative Tasks
7. Diffusion models etc.

LLMs super-charged next word predictors



AI in your Wavestone projects

Identify, clearly the functionality it should offer

Do you really need AI for this?

If so, what data will be used ?

Which type of AI would be best?

Are you happy with the ethical issues?