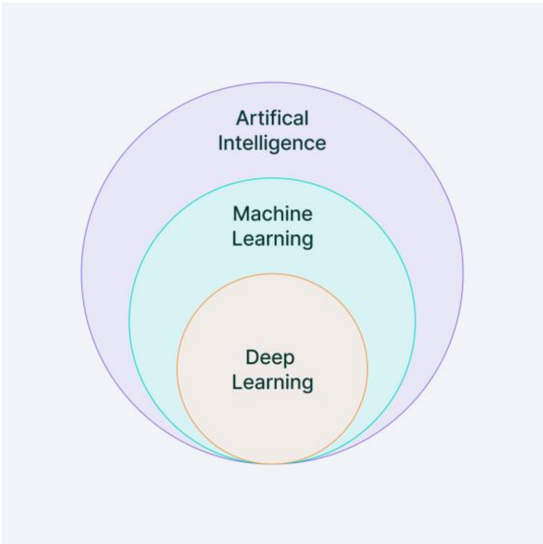
Machine Learning

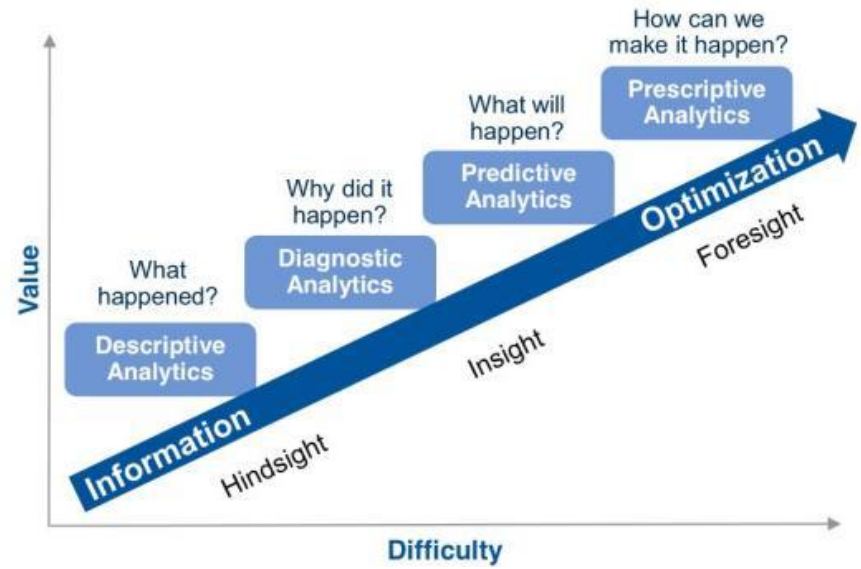
A comprehensive overview

Marcel van Velzen
Junior Marte Garcia

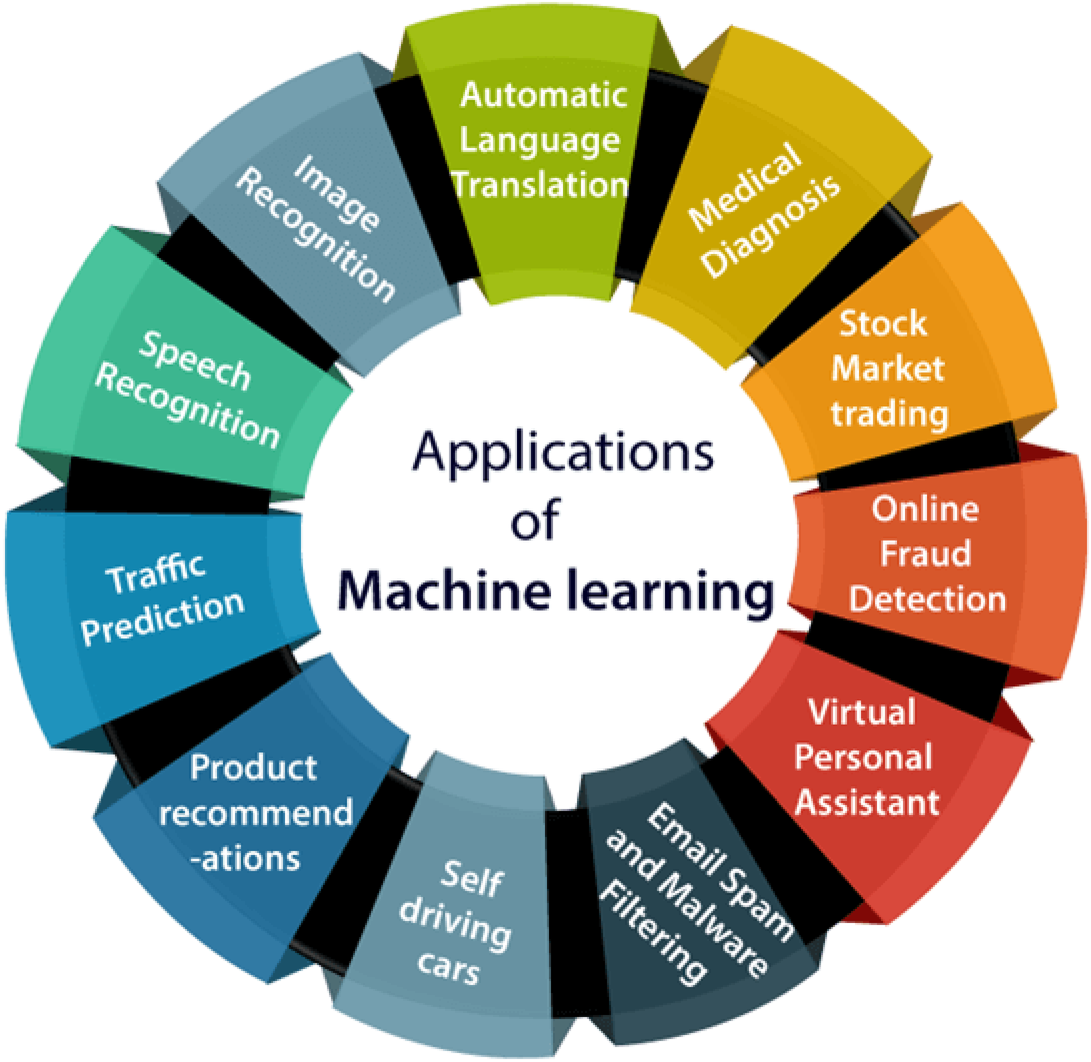
Definition



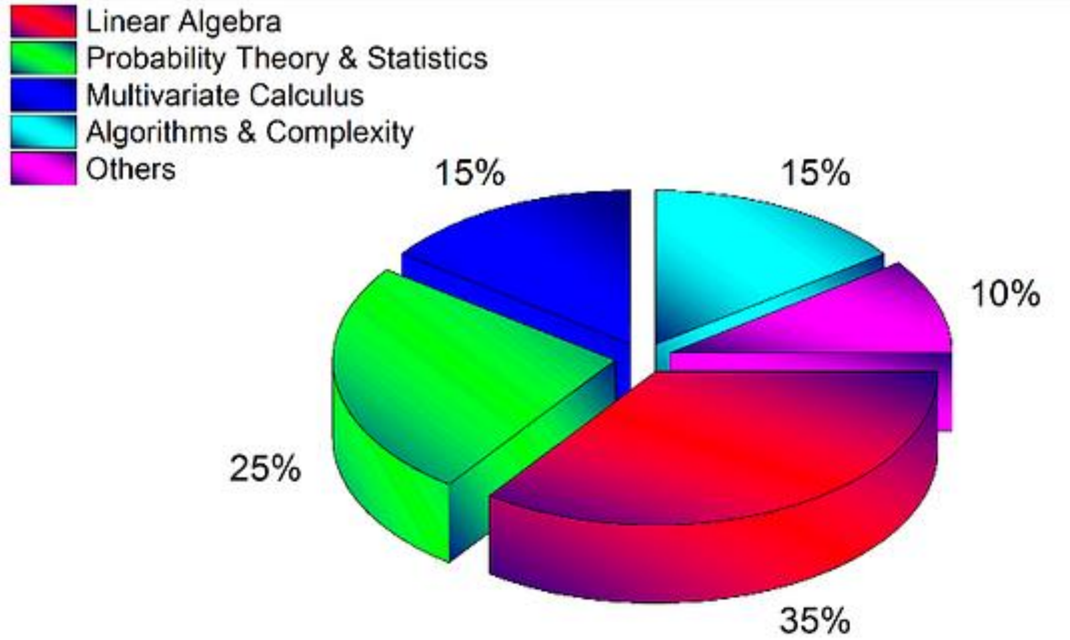
Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.



Applications



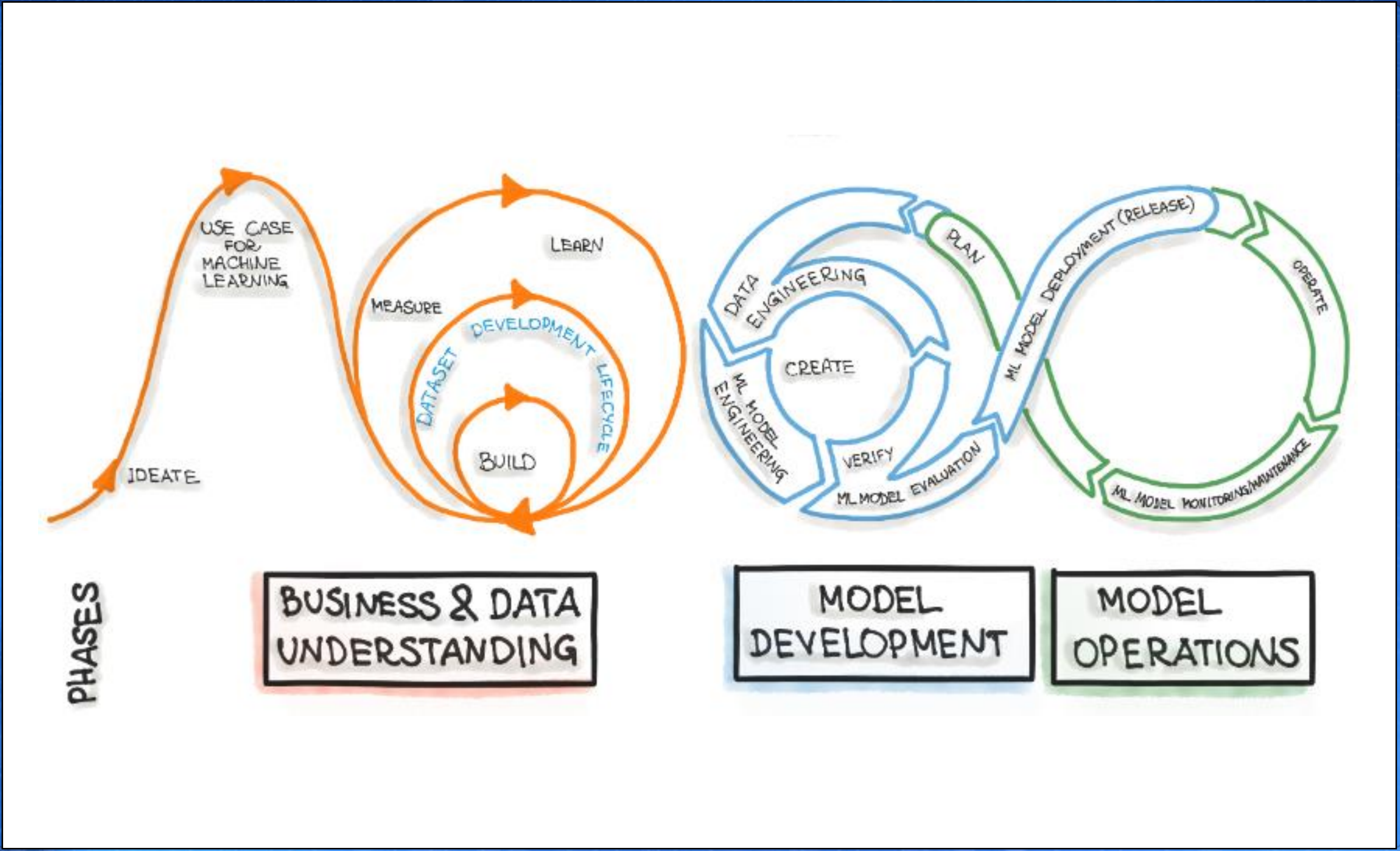
Math topics



Importance of Maths Topics Needed For Machine Learning

ML Lifecycle process

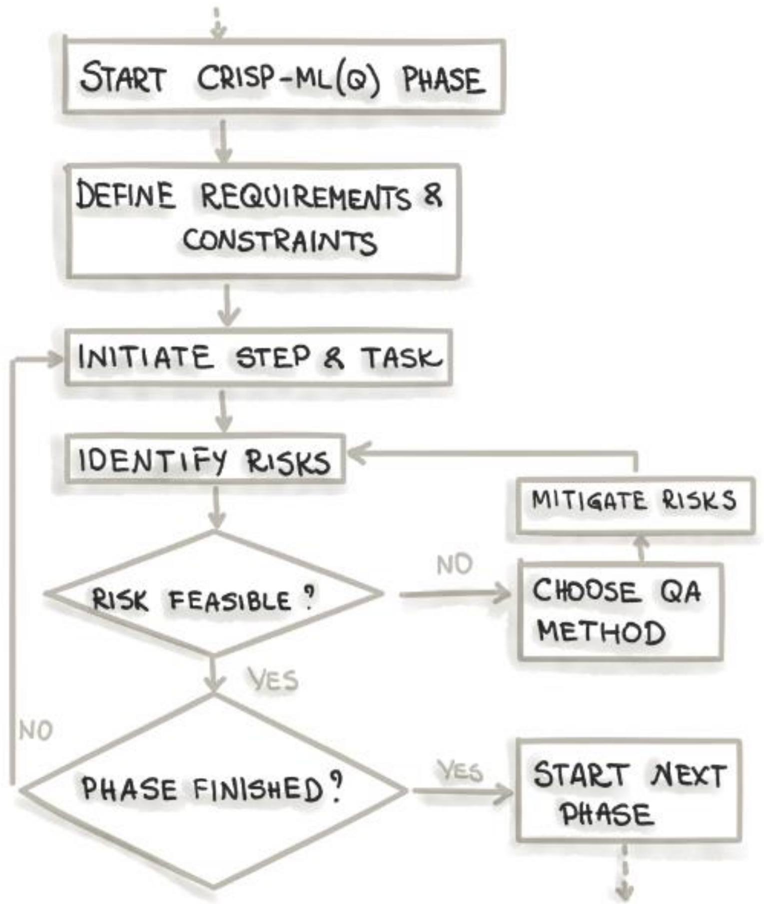
CRISP-ML(Q)



USE CASE FOR MACHINE LEARNING
IDEATE
MEASURE
LEARN
DATASET
DEVELOPMENT
LIFECYCLE
BUILD
DATA ENGINEERING
CREATE
ML MODEL ENGINEERING
VERIFY
ML MODEL EVALUATION
PLAN
ML MODEL DEPLOYMENT (RELEASE)
OPERATE
ML MODEL MONITORING/MAINTENANCE
PHASES
BUSINESS & DATA UNDERSTANDING
MODEL DEVELOPMENT
MODEL OPERATIONS

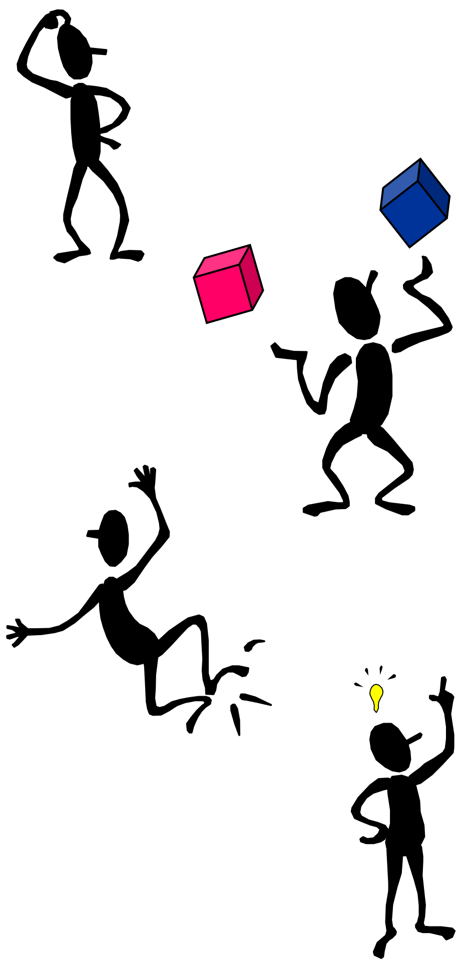
Phases

- Business and Data Understanding
- Data Engineering (Data Preparation)
- Machine Learning Model Engineering
- Quality Assurance for Machine Learning Applications
- Deployment
- Monitoring and Maintenance



Tasks

Business and Data Understanding	- Define business objectives - Translate business objectives into ML objectives - Collect and verify data - Assess the project feasibility - Create POC
Data Engineering	- Feature selection - Data selection - Class balancing - Cleaning data (noise reduction, data imputation) - Feature engineering (data construction) - Data augmentation - Data standartization
ML Model Engineering	- Define quality measure of the model - ML algorithm selection (baseline selection) - Adding domain knowledge to specialize the model - Model training - Optional: applying trainsfer learning (using pre-trained models) - Model compression - Ensemble learning - Documenting the ML model and experiments
ML Model Evaluation	- Validate model's performance - Determine robustess - Increase model's explainability - Make a decision whether to deploy the model - Document the evaluation phase
Model Deployment	- Evaluate model under production condition - Assure user acceptance and usability - Model governance - Deploy according to the selected strategy (A/B testing, multi-armed bandits)
Model Monitoring and Maintenance	- Monitor the efficiency and efficacy of the model prediction serving - Compare to the previously specified success criteria (thresholds) - Retrain model if required - Collect new data - Perform labelling of the new data points - Repeat tasks from the *Model Engineering* and *Model Evaluation* phases - Continuous, integration, training, and deployment of the model



ML Data Engineering

Characteristics of data

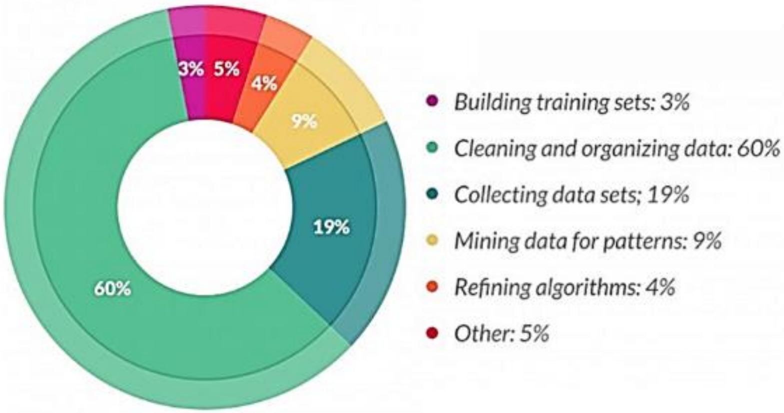
Data in the real world is dirty

- **Incomplete –** lacking attributes values, attributes of interest or only containing aggregate data
- **Noisy –** containing errors or outliers
- **Inconsistent –** containing discrepancies in codes or names

No quality data means no quality results

- Quality decisions can only be based on quality data

Tasks



- **Data duplication –** removal of duplicated observations
- **Dimensionality reduction –** find projection that captures observations
- **Encoding –** transform data so that it can be consumed by the modeling methods
- **Missing data –** delete, estimate, ignore and replace observations
- **Noise reduction –** remove unwanted or clean up polluted observations
- **Outliers –** identifying observations that do not fit the others
- **Feature selection –** remove redundant or irrelevant attributes
- **Sampling –** find a representative subset
- **Transformation –** map attribute values/objects into a single attribute value/object

Encoding

Nominal

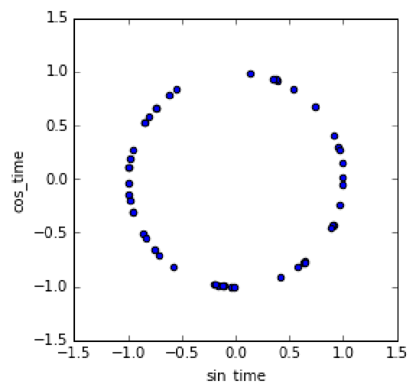
Color
Red
Red
Yellow
Green
Yellow

Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1
0	1	0

Ordinal

Original Encoding	Ordinal Encoding
Poor	1
Good	2
Very Good	3
Excellent	4

Cyclical continuous

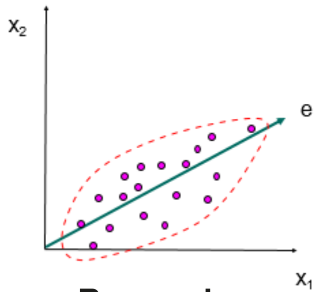


Data imputation

Missing data – data is not always available

- ❑ Ignore the subject
- ❑ Fill in the missing value manually or use a global constant
- ❑ Use the attribute mean to fill in the missing value
- ❑ Use the attribute mean for all subjects of the same class
- ❑ Predict the missing value with the most probable value

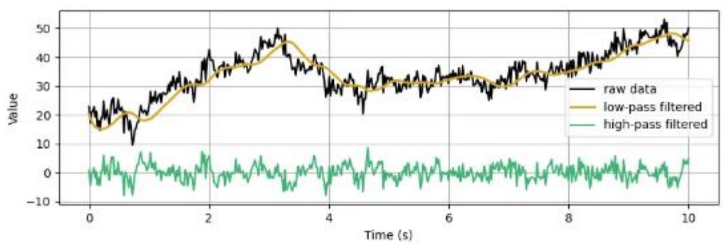
Noise reduction



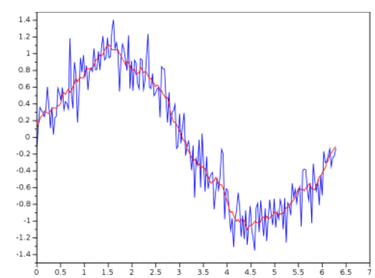
Regression

Age				
Category	<20	21-40	41-65	>65

Binning



Filters



Moving averages

Signal processing

	Smoothness	Responsive	Score
KAMA	5	6	11
VIDYA	6	5	11
MAMA	7	1	8
Ehlers	1	3	4
Median	8	7	15
Median-MA	4	8	12
FRAMA	2	4	6
Laguerre	3	2	5

Filters are designed to selectively modify or extract specific frequency components from a signal while attenuating others

Low-pass filter: Allows low-frequency components to pass through while attenuating higher frequencies. It is useful for removing high-frequency noise or extracting the slow-changing trends from a signal.

High-pass filter: Allows high-frequency components to pass through while attenuating lower frequencies. It is used to remove low-frequency noise or isolate fast-changing features in a signal.

Band-pass filter: Allows a specific range of frequencies to pass through while attenuating others. It is employed when you want to isolate a specific band of frequencies from a signal.

Notch filter: Attenuates a narrow band of frequencies, often used to remove specific interference or noise components.

Different filter designs may be suitable for different scenarios.

Moving averages

$$SMA_n = \frac{\sum_{i=1}^{n} close_i}{n}$$

$$WMA_{t+1} = \frac{\sum_{i=t-n}^{t}(x_i * w_i)}{\sum_{i=1}^{n} w_i}$$

$$EMA_n = \alpha * close_n + (1-\alpha) * EMA_{n-1}, \alpha = \frac{2}{n+1}$$

$$\text{Triple Exponential Moving Average (TEMA)} = (3 * EMA_1) - (3 * EMA_2) + EMA_3$$

Moving averages are a form of smoothing technique that calculates the average of a subset of adjacent data points within a time series. A "window" or "kernel" of fixed size moves along the data, and the average value within the window is computed at each position.

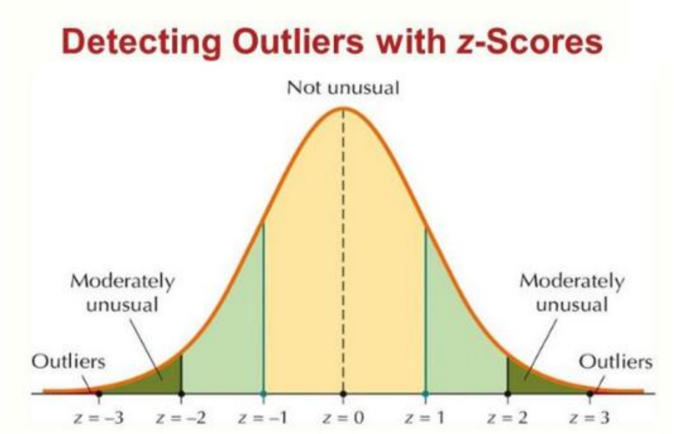
Noise reduction: By averaging out nearby data points, moving averages can suppress high-frequency noise or fluctuations, resulting in a smoother representation of the underlying signal.

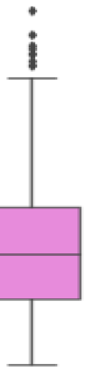
Trend and momentum analysis: Moving averages can help identify trends, momentum and patterns in data by smoothing out short- and longer-term fluctuations. Different types of moving averages, such as simple moving averages (SMA), weighted moving averages (WMA), and exponential moving averages (EMA), provide varying emphasis on recent versus older data points.

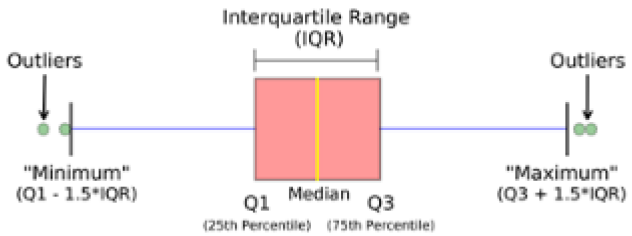
Forecasting: Moving averages can be used to generate predictions or forecasts by extrapolating the smoothed trend. They are often employed in time series analysis and financial markets to make short-term predictions.

Different window sizes may be suitable for different scenarios.

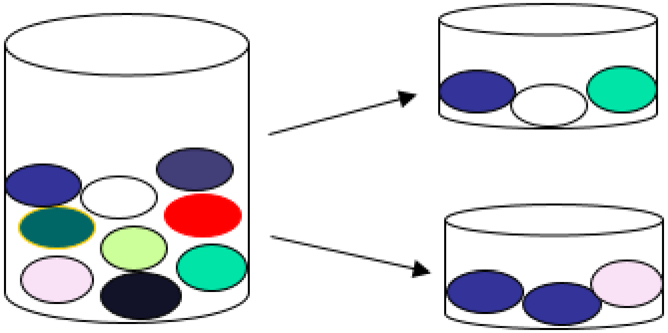
Outlier detection



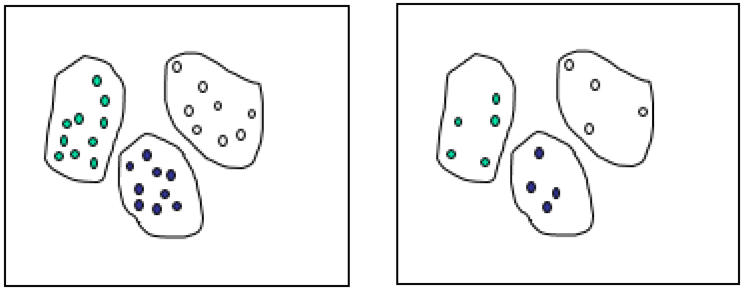




Sampling

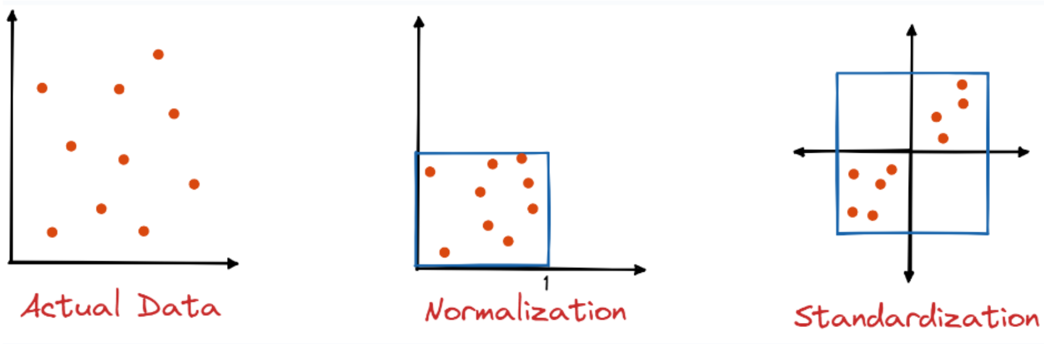


Random sampling



Stratified sampling

Normalization and Standardization



$$z = \frac{\mathrm{x} - min(x)}{[\max(x) - min(x)]}$$

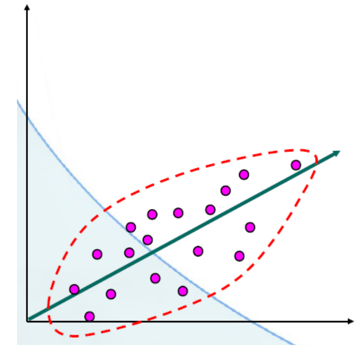
$$z = \frac{\mathrm{x} - mean(x)}{stdev(x)}$$

$$z = \frac{1}{1 + \exp(-x)}$$

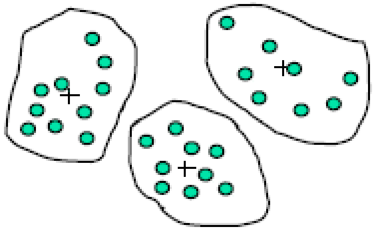
ML Model Engineering

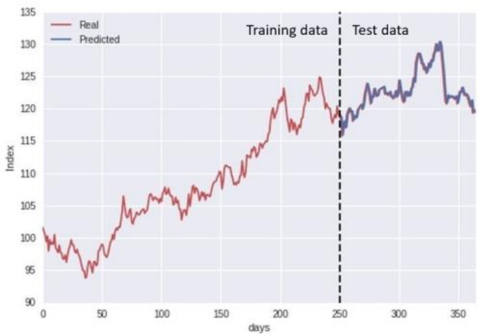
Supervised Learning

Regression – assessing the relationship between traits in the data and calculates how much a variable changes when any other variables change



Classification – identify new occurrences which category it belongs to based on known observations

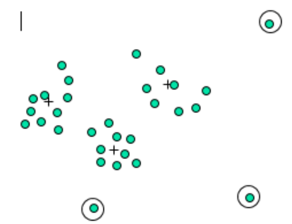




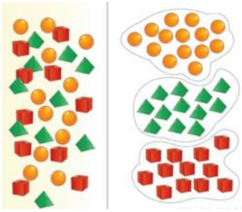
Time series – predicting future values based on historical data

Unsupervised Learning

Outlier detection – detecting occurrences which fall outside the norm

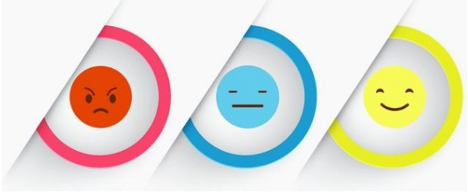


Association rules – searching and identifying dependencies, relationships or orders between features in the data



Clustering – grouping occurrences which have common properties

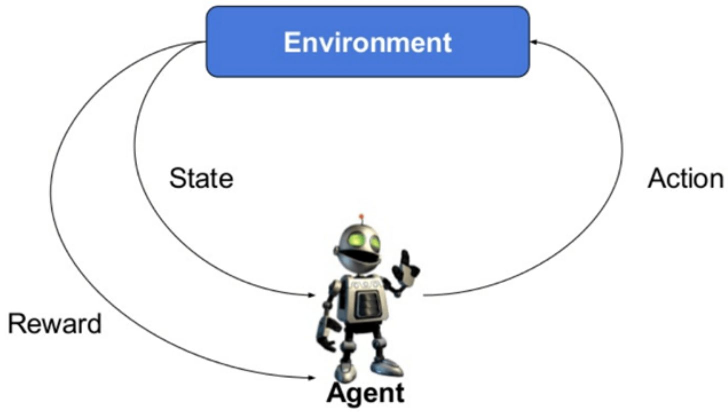
Sentiment analysis – measuring of the sentiment

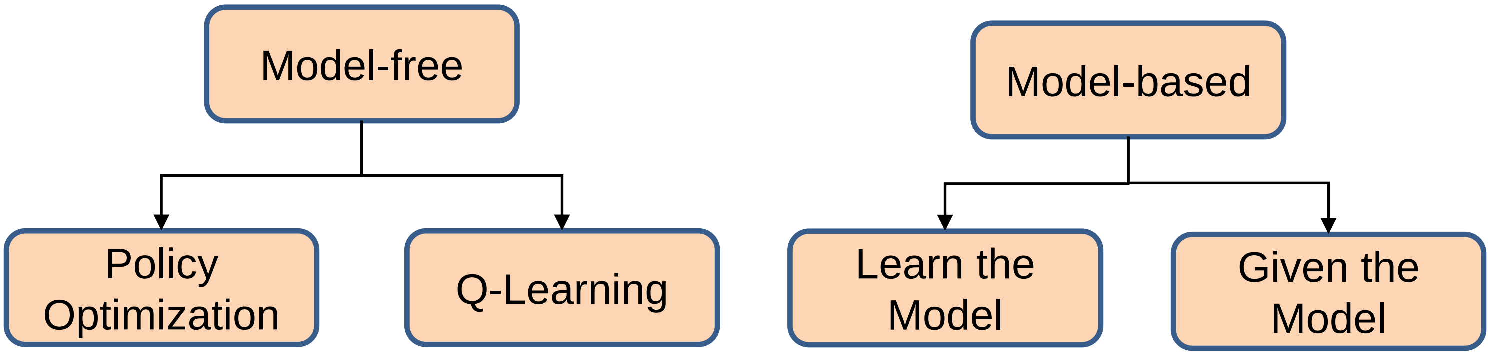


Dimensionality reduction – compresses the data that encapsulates the original data

Reinforcement Learning

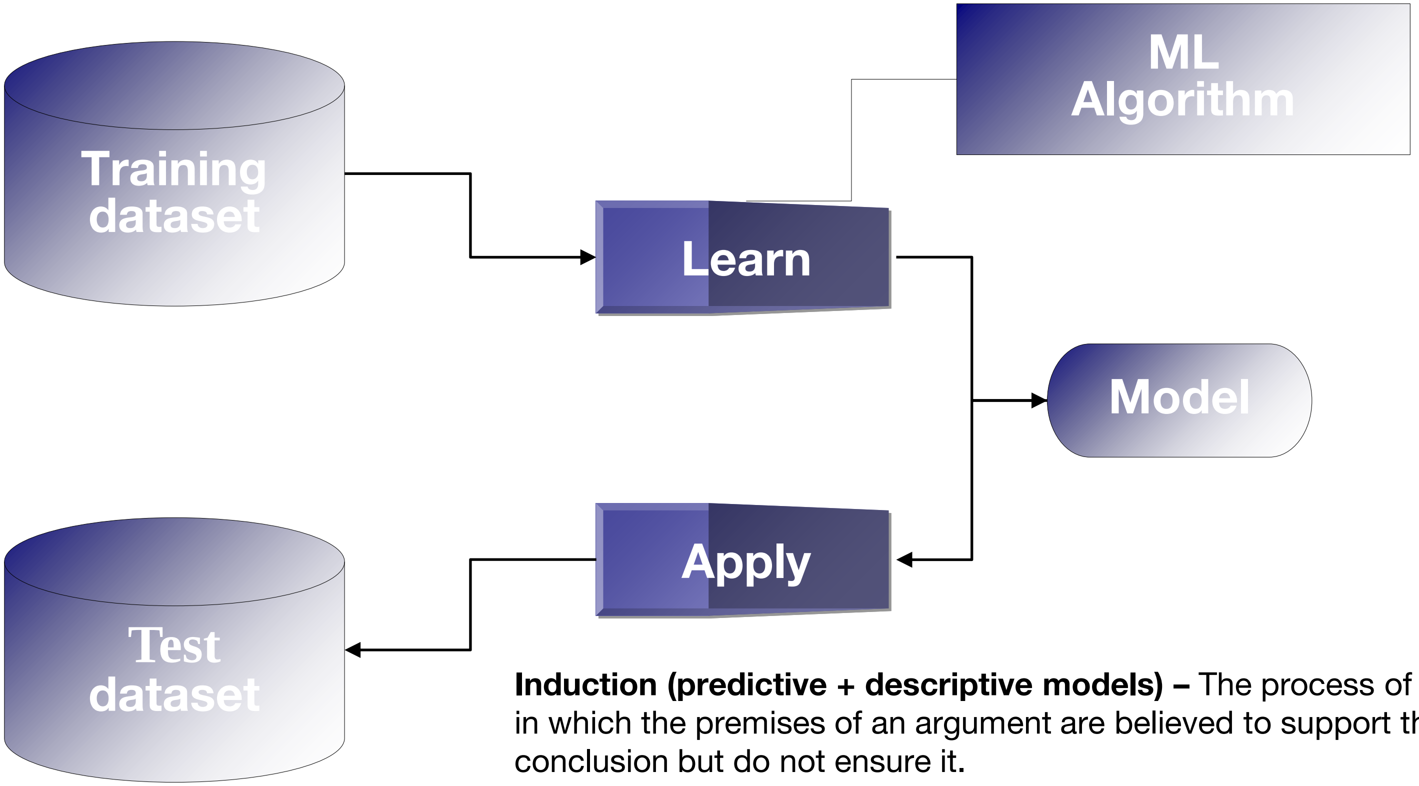
Typical RL scenario





Learning

The analysis of data for discovering meaningful new correlations, patterns and trends by building models (representations of reality) using machine learning (statistical or artificial intelligence) techniques.

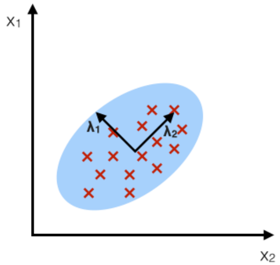


Induction (predictive + descriptive models) – The process of reasoning in which the premises of an argument are believed to support the conclusion but do not ensure it.
Deduction (only predictive models) – The kind of reasoning in which the conclusion is necessitated by, or reached from, previously known facts (the premises)

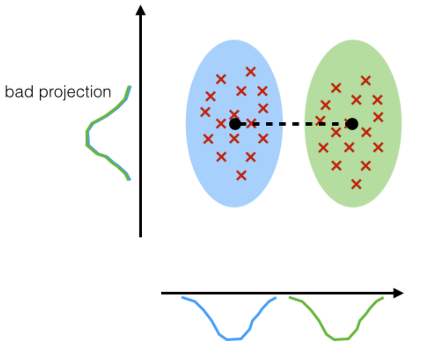
ML Algorithms

Dimensionality reduction

PCA:
component axes that maximize the variance



LDA:
maximizing the component axes for class-separation

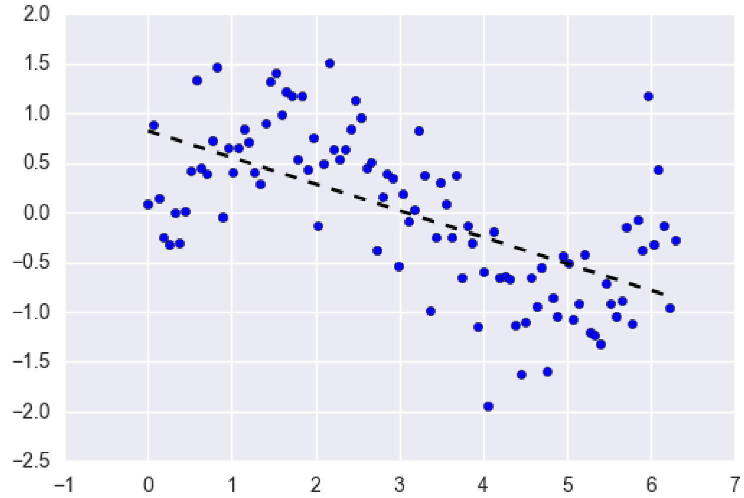


Principal Component Analysis (PCA) is an unsupervised learning technique that aims to maximize the variance of the data along the principal components

Linear discriminant analysis (LDA) is a supervised learning technique that aims to maximize the separation between different classes in the data

- ❑ PCA is unsupervised and focuses on maximizing variance, while LDA is supervised and focuses on maximizing class separation
- ❑ PCA doesn't require labeled data, but LDA does
- ❑ PCA reduces dimensionality by projecting data onto a lower-dimensional space, while LDA creates linear combinations of features
- ❑ PCA outputs principal components that capture variation, while LDA outputs discriminant functions that separate classes
- ❑ PCA is commonly used for exploratory data analysis, while LDA is often used for classification tasks
- ❑ PCA is generally faster and more computationally efficient, but LDA may be more effective with labeled data

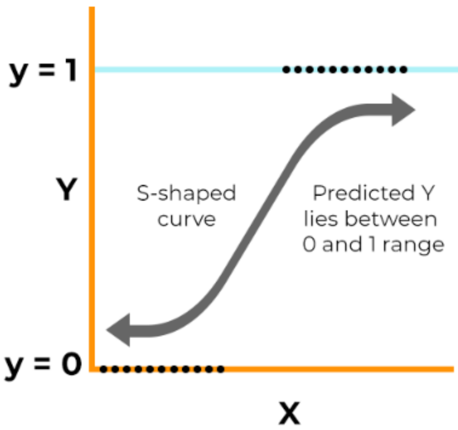
Linear Regression



Linear regression is a statistical technique that models the relationship between a dependent variable and one or more independent variables by fitting a straight line to the data

Benefits	Drawbacks
works well when there are linear relationships between the variables in your dataset	performs poorly when there are non-linear relationships
straightforward to understand and explain	often outclassed by its regularized counterparts
can be updated easily with new data	

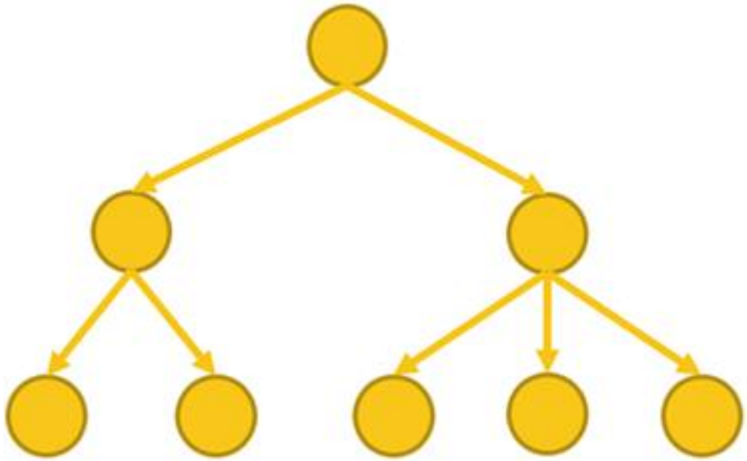
Logistic Regression



Logistic regression is a statistical technique used to model the relationship between a binary dependent variable and one or more independent variables by estimating the probability of the dependent variable belonging to a certain category

Benefits	Drawbacks
easy to implement, interpret, and efficient in training	may lead to overfitting if the number of observations is fewer than the number of features
flexible and does not assume specific class distributions	constructs linear boundaries and assumes linearity between the dependent and independent variables
can handle multiple classes and provides a probabilistic view of predictions	limited to predicting discrete functions and is not suitable for non-linear problems
measures predictor importance and direction of association	requires low or no multicollinearity among independent variables
quickly classifies unknown records and performs well with linearly separable datasets	may struggle to capture complex relationships

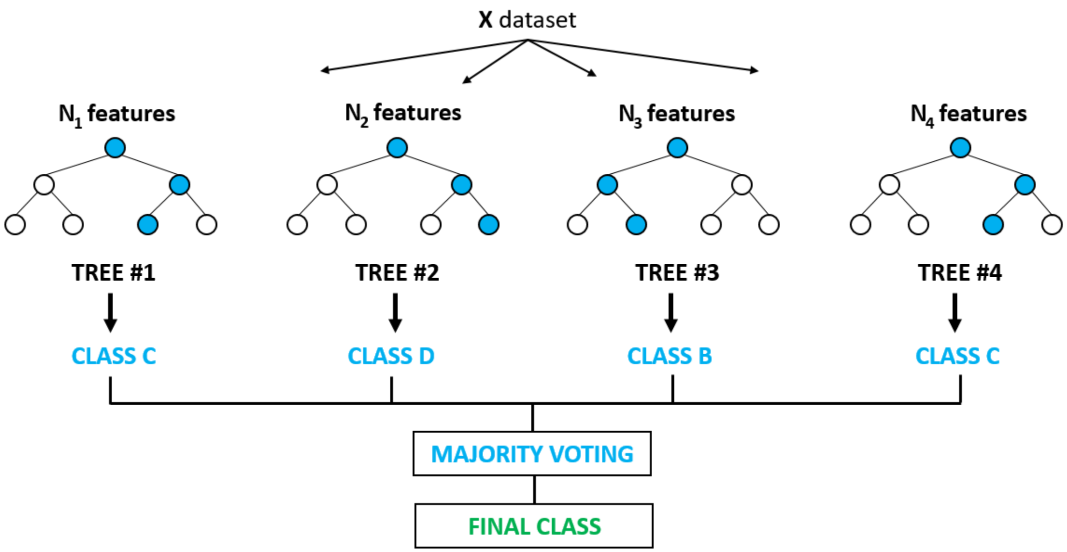
Decision Tree



A **decision tree** in machine learning is a hierarchical model that uses a sequence of binary splits based on input features to make predictions or classify data

Benefits	Drawbacks
require less effort for data preparation during pre-processing compared to other algorithms	prone to instability
normalization of data is not necessary	calculation more complex
scaling of data is not required	higher training time
missing values in the data have minimal impact	relatively expensive
decision tree is intuitive and easy to explain	primarily designed for predicting discrete or categorical values rather than continuous values

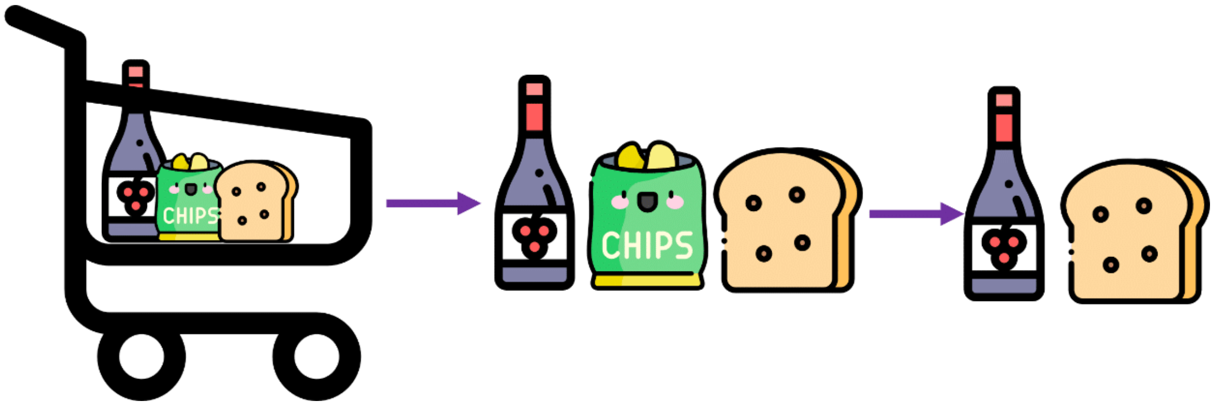
Random Forest



Random Forest combines the output of multiple decision trees to reach a single result to make predictions or classify data

Benefits	Drawbacks
Versatile and easy to use	Computationally demanding
Handles high-dimensional spaces	Model interpretability
Feature importance	Overcomplexity
Robust to overfitting	Bias in multiclass problems
Out-of-box predictor	Lack of precision

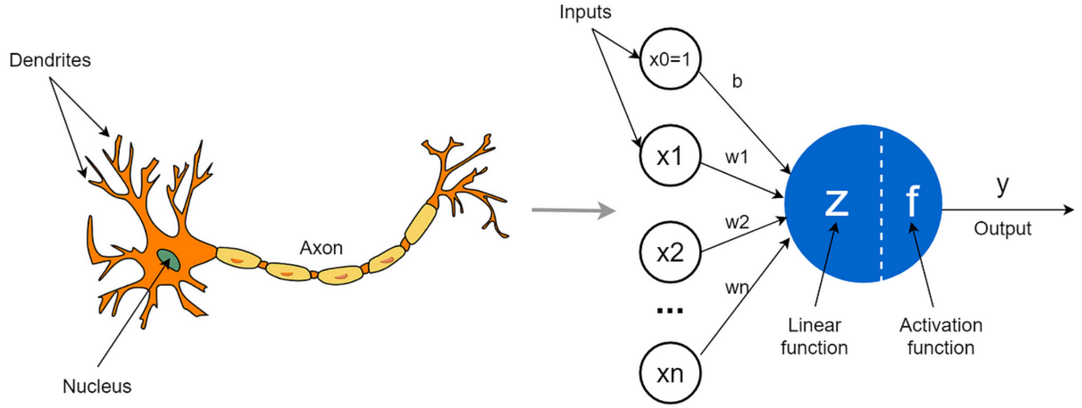
Apriori algorithm



The **Apriori algorithm** finds frequent patterns and associations in a transactional dataset

Benefits	Drawbacks
Simplicity and ease of implementation	Computational complexity
The rules are human-readable	Difficulty handling sparse data
Flexible and customisable	Limited discovery of complex patterns
The algorithm is widely used and studied	Bias of minimum support threshold
	Inability to handle numeric data

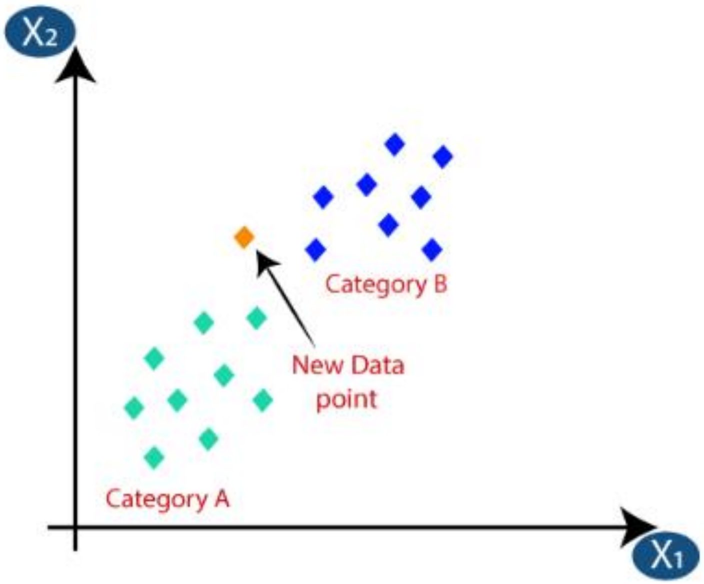
Neural network



A **neural network** in machine learning is a computational model inspired by the structure and function of the human brain, composed of interconnected nodes or "neurons" that process and transmit information to make predictions or classify data

Benefits	Drawbacks
have several advantages over traditional algorithms	complex and require a significant amount of data to train
can learn from data and tackle complex problems	overfitting is a concern
can generalize and identify patterns that traditional algorithms may miss	lack interpretability
particularly useful for tasks like image recognition and natural language processing	less suited for reasoning or decision-making
are efficient at processing large amounts of data with speed and accuracy	lack explanatory capabilities

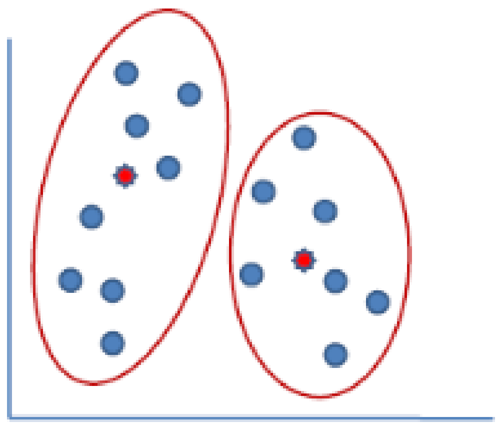
K-Nearest Neighbor



The **k-nearest neighbor** algorithm is a machine learning algorithm that classifies new data points based on the majority vote of their k nearest neighbors in the training data

Benefits	Drawbacks
simple to implement	needs to determine the value of k
robust to the noisy training data	computation cost is high
can be more effective if the training data is large	

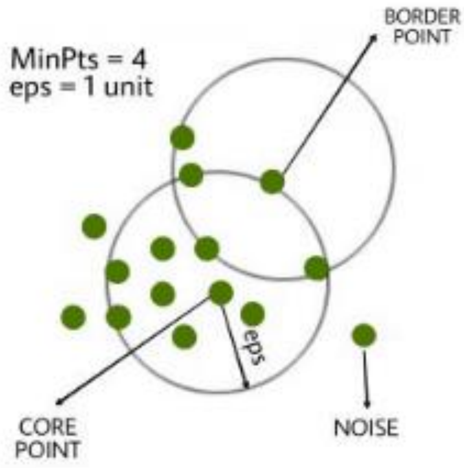
K-Means



The **K-means** algorithm is a clustering technique that aims to partition data points into k distinct clusters based on their similarity, where each data point is assigned to the cluster with the nearest mean value

Benefits	Drawbacks
relatively easy to implement and apply	determining the optimal value of k
can handle large datasets effectively	dependence on initial values can impact the results of k-means clustering
guarantees convergence to a final solution	clustering data with varying sizes and density can be challenging
allows for warm-starting, initializing centroids with predefined positions	outliers can affect the clustering results
can easily adapt to new examples and generalize to clusters of different shapes and sizes	scalability of k-means is influenced by the number of dimensions in the data

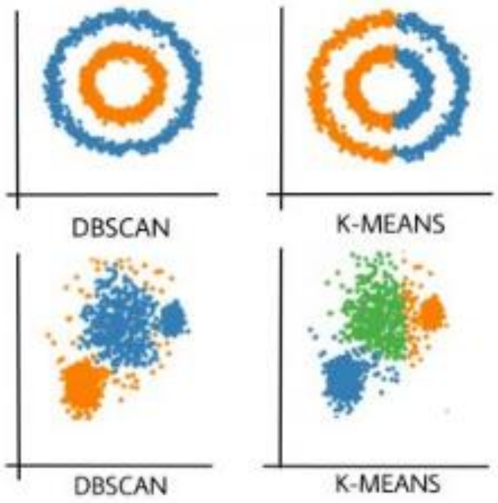
DBSCAN



The **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** algorithm is a density-based clustering algorithm that groups data points based on their density and identifies outliers as noise.

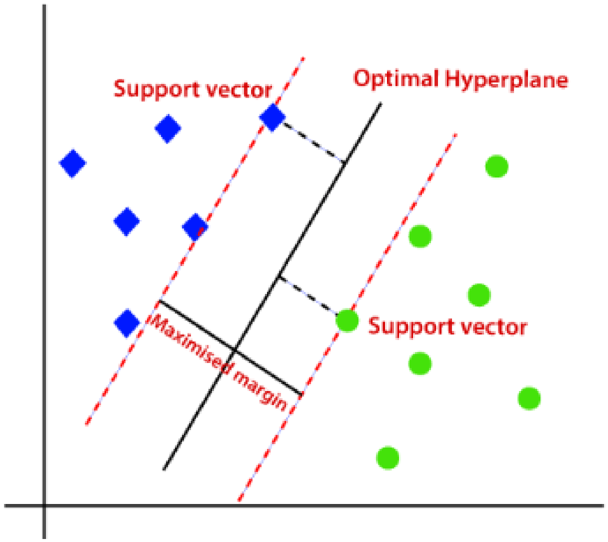
Benefits	Drawbacks
Handles irregularly shaped and sized clusters	Not suitable for datasets with categorical features
Robust to outliers	Requires a drop in density to detect cluster borders
Does not require the number of clusters to be specified	Struggles with clusters of varying density Sensitive to scale of variables
Less sensitive to initialization conditions	Sensitive to scale of variables
Relatively fast compared to other clustering algorithms	Performance tends to degrade in high-dimensional data

Difference DBSCAN and K-Means



DBSCAN	K-Means
In DBSCAN we need not specify the number of clusters	K-Means is very sensitive to the number of clusters so it need to be specified
Clusters formed in DBSCAN can be of any arbitrary shape	Clusters formed in K-Means are spherical or convex in shape
DBSCAN can work well with datasets having noise and outliers	K-Means does not work well with outliers data, ouliers can skew the clusters in K-Means to a very large extent
In DBSCAN two parameters are required for training the model	In K-Means only one parameter is required for training the model

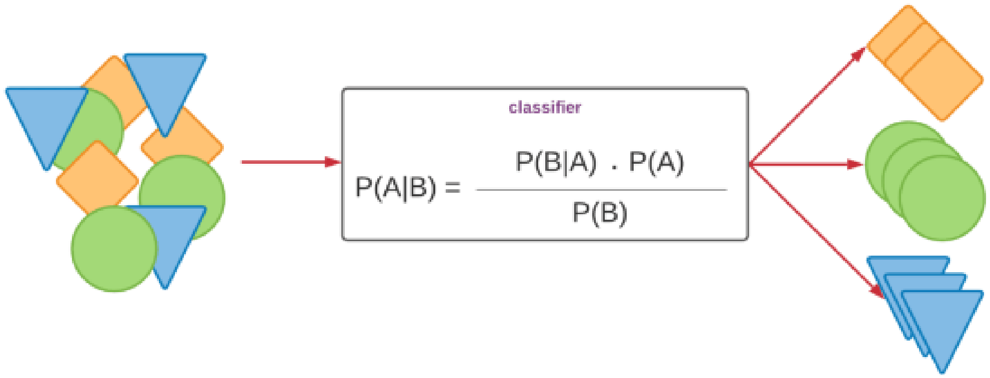
Support Vector Machine



The **Support Vector Machine (SVM)** algorithm is a supervised learning algorithm that separates data points by finding the optimal hyperplane with the largest margin between different classes

Benefits	Drawbacks
works better when the data is linear	choosing a good kernel is not easy
more effective in high dimensions	doesn’t show good results on a big dataset
can solve any complex problem with kernel trick	not that easy to fine-tune the hyper-parameters
not sensitive to outliers	
can do image classifications	

Naive Bayes

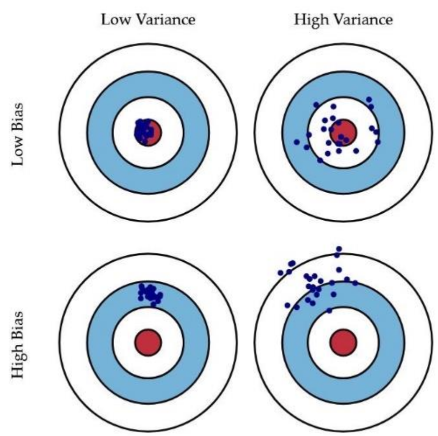


The **Naive Bayes** algorithm is a simple probabilistic classifier that calculates the probability of a data point belonging to a certain class based on the conditional probabilities of its features

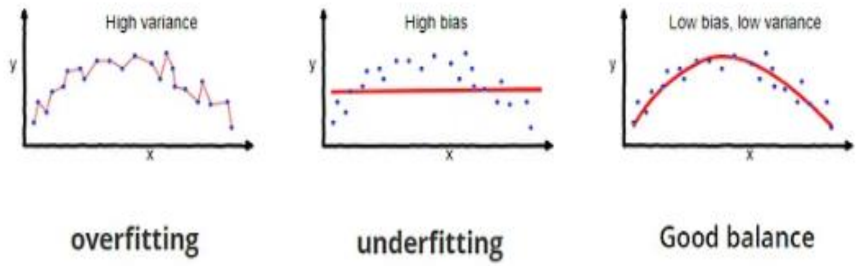
Benefits	Drawbacks
works quickly and can save a lot of time	assumes that all predictors (or features) are independent
suitable for solving multi-class prediction problems	faces the ‘zero-frequency problem’
can perform better than other models and requires much less training data	estimations can be wrong in some cases
better suited for categorical input variables than numerical variables	

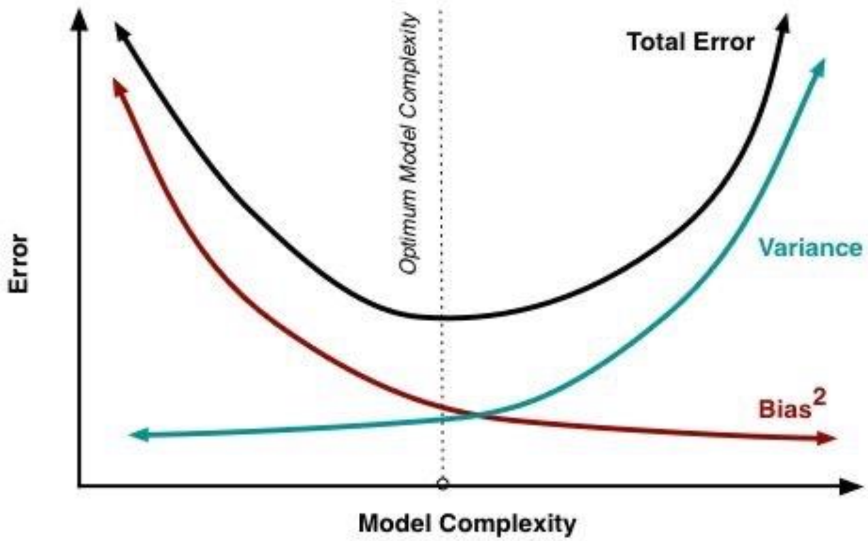
ML Model Evaluation

Bias-Variance Tradeoff



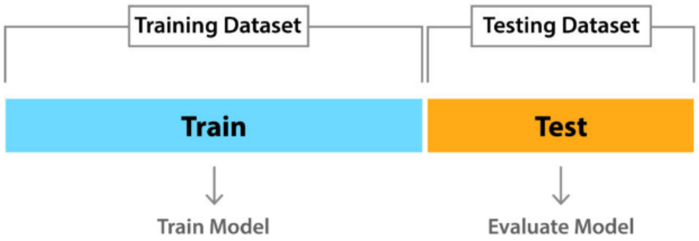
There is a tradeoff between a model's ability to minimize bias and variance





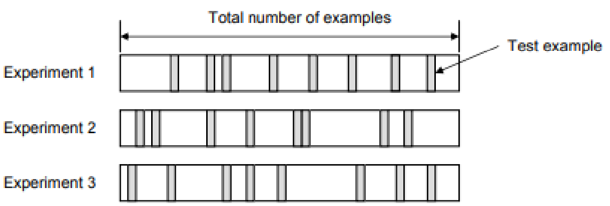
Cross Validation

Hold out



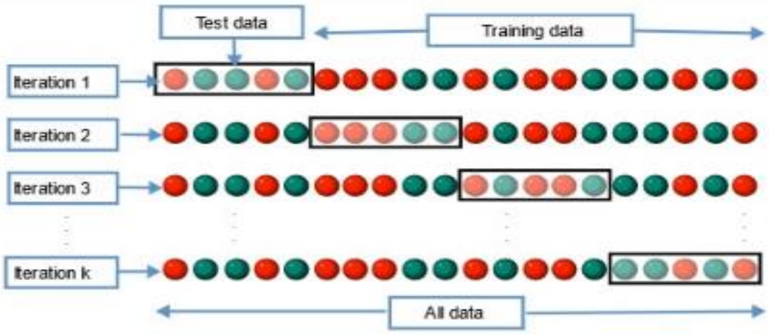
- ❑ Simple and easy to use
- ❑ Not enough test data for a sparse dataset
- ❑ Error rate misleading when unfortunate split

Random subsampling



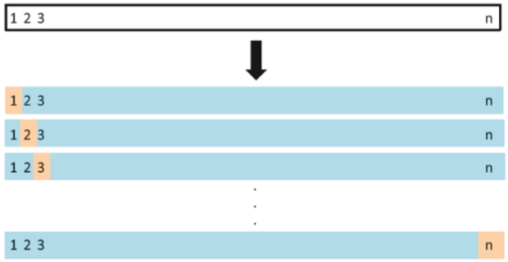
- ❑ simplicity and lack of bias
- ❑ larger population necessary
- ❑ under certain circumstances bias can occur

K-Fold



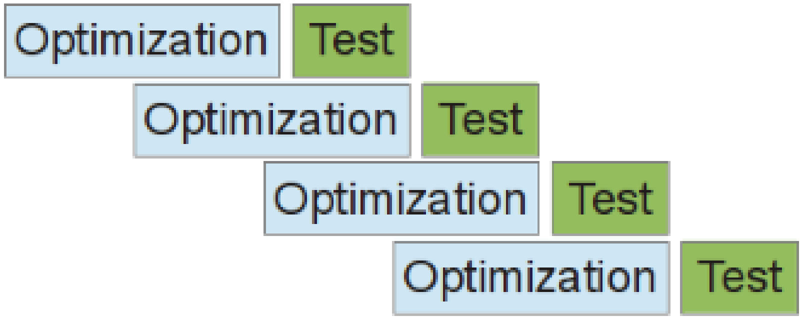
- ❑ stable accuracy
- ❑ prevents overfitting
- ❑ model Generalization Validation
- ❑ imbalanced dataset
- ❑ computational costs

Leave one out

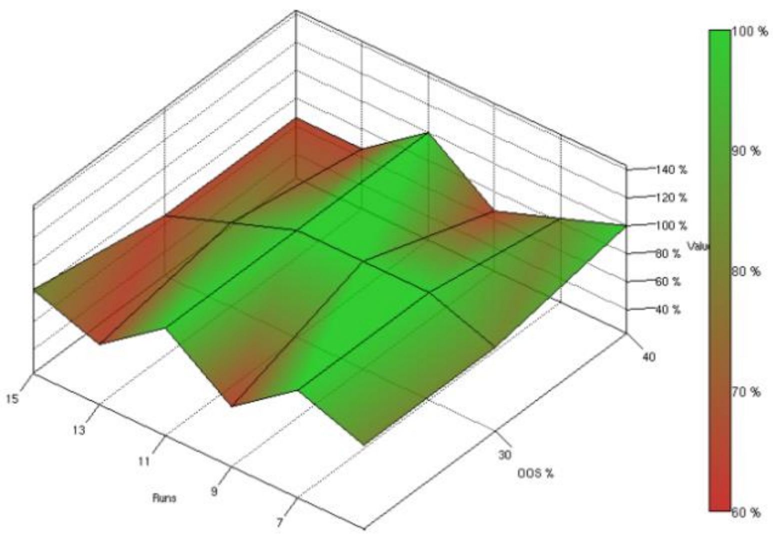


- ❑ less biased on test data
- ❑ computational costs

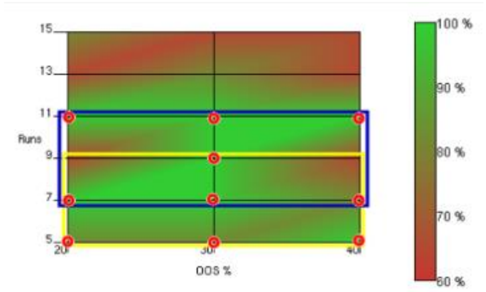
Walk Forward Optimization



Walk Forward Analysis does optimization on a training set; test on a period after the set and then rolls it all forward and repeats the process.



Walk Forward Matrix is a set of walk forward analysis with different number of periods and out of sample percentages



Regression Metrics

$$MSE = \frac{1}{N} \times \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

N - number of data samples
y_i - actual data value
$\hat{y}_i$ - predicted data value

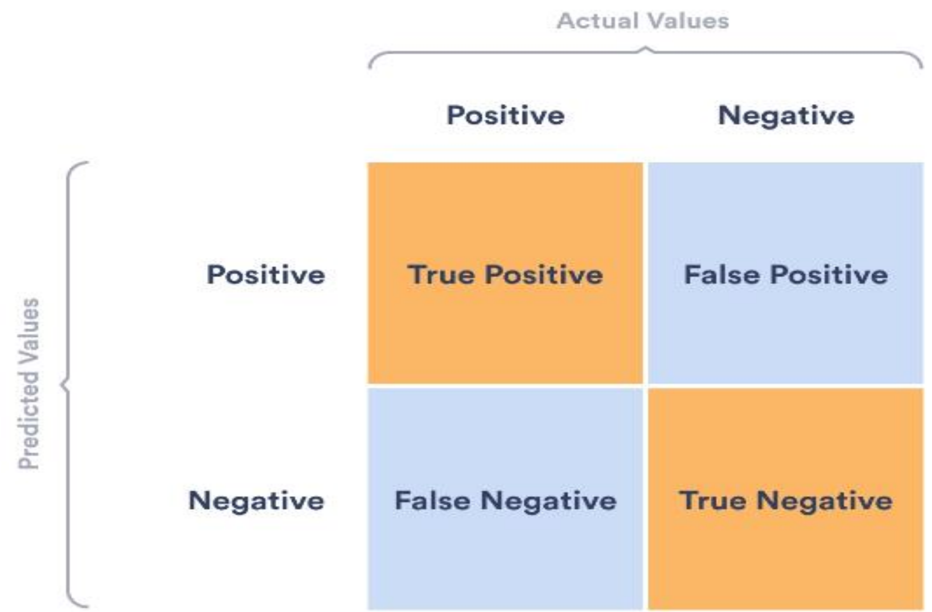
$$RMSE = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

N - number of data samples
y_i - actual data value
$\hat{y}_i$ - predicted data value

$$MAE = \frac{1}{N} \times \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

N - number of data samples
y_i - actual data value
$\hat{y}_i$ - predicted data value

Classification Metrics



A confusion matrix visualizes and summarizes the performance of a classification algorithm

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Precision} = \frac{\text{Number of correct positive results}}{\text{Total number of positive results}} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{Number of correct positives}}{\text{Number of all positives}} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Specificity} = \frac{\text{Number of correctly predicted negatives}}{\text{Number of all negatives}} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

$$\text{F1 Score} = \text{Harmonic mean of Precision and Recall} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2\,TP}{2TP + FP + FN}$$

Associations Metrics

Transaction 1	🍎 🍺 🍚 🍗
Transaction 2	🍎 🍺 🍚
Transaction 3	🍎 🍺
Transaction 4	🍎 🍐
Transaction 5	🍼 🍺 🍚 🍗
Transaction 6	🍼 🍺 🍚
Transaction 7	🍼 🍺
Transaction 8	🍼 🍐

Support {🍎}

This says how popular an itemset is, as measured by the proportion of transactions in which an itemset appears
The support of {apple} is 4 out of 8, or 50%

Confidence {🍎 → 🍺} = Support {🍎, 🍺} / Support {🍎}

This says how likely item Y is purchased when item X is purchased, expressed as {X -> Y}
The confidence of {apple -> beer} is 3 out of 4, or 75%

Lift {🍎 → 🍺} = Support {🍎, 🍺} / (Support {🍎} x Support {🍺})

This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is
The lift of {apple -> beer} is 3 out of 4 multiplied by 6, or 12,5%

References

Akinfaderin, W. (2021, April 30). The Mathematics of Machine Learning - Towards Data Science. Medium. https://towardsdatascience.com/the-mathematics-of-machine-learning-894f046c568

AlgoDaily. (n.d.). AlgoDaily - Daily coding interview questions. Full programming interview prep course and software career coaching. https://algodaily.com/lessons/standardization-and-normalization

Anindya. (2022). Naive Bayes algorithm in Machine Learning with Python. ThinkInfi. https://thinkinfi.com/naive-bayes-algorithm-in-machine-learning-with-python/

Applications of Machine Learning - Javatpoint. (n.d.). www.javatpoint.com. https://www.javatpoint.com/applications-of-machine-learning

Baheti, P. (2023, April 24). What is Machine Learning? The Ultimate Beginner's Guide. V7. https://www.v7labs.com/blog/machine-learning-guide

Butler, R. G. (2017, May 22). Preparing your Data for AutoDiscovery: Table-Like Structure in Excel. butlerscientifics. https://www.butlerscientifics.com/single-post/2017/05/22/preparing-your-data-for-autodiscovery-table-like-structure-in-excel

Chauhan, A. (2021, December 31). Random Forest Classifier and its Hyperparameters - Analytics Vidhya - Medium. Medium. https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6

Descriptive Predictive Prescriptive Analytics | Data Science Association. (n.d.). https://www.datascienceassn.org/content/descriptive-predictive-prescriptive-analytics

Editor. (2021, October 30). Data Scientist vs Data Engineer: Differences and Why You Need Both. AltexSoft. https://www.altexsoft.com/blog/data-scientist-vs-data-engineer/

Ehlers, J. (2005), Building Trading Systems on Nonlinear Filters

EliteDataScience. (2022, July 8). Modern Machine Learning Algorithms: Strengths and Weaknesses. EliteDataScience. https://elitedatascience.com/machine-learning-algorithms

Ellis, C. (2022, June 8). When to use DBSCAN. Crunching the Data. https://crunchingthedata.com/when-to-use-dbscan/

Encoding cyclical continuous features - 24-hour time. (2016, July 31). Ian London's Blog. https://ianlondon.github.io/blog/encoding-cyclical-features-24hour-time/

GeeksforGeeks. (2023). Advantages and Disadvantages of Logistic Regression. GeeksforGeeks. https://www.geeksforgeeks.org/advantages-and-disadvantages-of-logistic-regression/

ChatGPT

References

GeeksforGeeks. (2023b). DBSCAN Clustering in ML   Density based clustering. GeeksforGeeks. https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/

Goyal, C. (2021). Importance of Cross Validation: Are Evaluation Metrics enough? Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/05/importance-of-cross-validation-are-evaluation-metrics-enough/

Han, J., Kamber, M., & Pei, J. (2012). Data mining: concepts and techniques. Choice Reviews Online, 49(06), 49–3305. https://doi.org/10.5860/choice.49-3305

K, D. (2023, February 7). Top 5 advantages and disadvantages of Decision Tree Algorithm. Medium. https://dhirajkumarblog.medium.com/top-5-advantages-and-disadvantages-of-decision-tree-algorithm-428ebd199d9a

k-Means Advantages and Disadvantages. (n.d.). Google for Develop

K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint. (n.d.). www.javatpoint.com. https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning

Kumar, A. (2023, April 13). PCA vs LDA Differences, Plots, Examples - Data Analytics. Data Analytics. https://vitalflux.com/pca-vs-lda-differences-plots-examples/#:~:text=PCA%20is%20an%20unsupervised%20learning,directions%20of%20maximum%20class%20separability

Kumar, A. (n.d.). AN INTRODUCTION TO MARKET BASKET ANALYSIS - ASSOCIATION RULE. www.linkedin.com. https://www.linkedin.com/pulse/introduction-market-basket-analysis-association-rule-abhishek-kumar

Lecture 12: Bias Variance Tradeoff. (n.d.). https://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote12.html

Likebupt. (2021, November 4). Normalize Data: Component Reference - Azure Machine Learning. Microsoft Learn. https://learn.microsoft.com/en-us/azure/machine-learning/component-reference/normalize-data?view=azureml-api-2

Mitchell, C. (2022). Triple Exponential Moving Average (TEMA): Definition and Formula. Investopedia. https://www.investopedia.com/terms/t/triple-exponential-moving-average.asp

ml-ops.org. (2023, February 22). https://ml-ops.org/content/crisp-ml

Ohseokkim. (2021). [Preprocessing] Encoding Categorical Data. Kaggle. https://www.kaggle.com/code/ohseokkim/preprocessing-encoding-categorical-data

Part 2: Kinds of RL Algorithms — Spinning Up documentation. (n.d.). https://spinningup.openai.com/en/latest/spinningup/rl_intro2.html

References

Pramoditha, R. (2022, January 3). The Concept of Artificial Neurons (Perceptrons) in Neural Networks. Medium. https://towardsdatascience.com/the-concept-of-artificial-neurons-perceptrons-in-neural-networks-fab22249cbfc

R, G. S. (2022, October 21). Encoding Categorical Data- The Right Way - Towards AI. Medium. https://pub.towardsai.net/encoding-categorical-data-the-right-way-4c2831a5755

Ricardo Gutierrez-Osuna, Lecture 13: Cross-validation, http://www.cs.tau.ac.il/~nin/Courses/NC05/pr_l13.pdf

RoboticsBiz. (2022). Pros and cons of Random Forest Algorithm. RoboticsBiz. https://roboticsbiz.com/pros-and-cons-of-random-forest-algorithm/

SagarDhandare. (2022, March 28). Nominal And Ordinal Encoding In Data Science! - Nerd For Tech - Medium. Medium. https://medium.com/nerd-for-tech/nominal-and-ordinal-encoding-in-data-science-c93872601f16

Saini, A. (2023). Support Vector Machine(SVM): A Complete guide for beginners. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/

Seraydarian, L. (2023). What Is a Confusion Matrix in Machine Learning? Plat.AI. https://plat.ai/blog/confusion-matrix-in-machine-learning/

Sumanth, G. (2021, December 12). Illustrative Example of Principal Component Analysis(PCA) vs Linear Discriminant Analysis(LDA): Is PCA good guy or bad guy ? Medium. https://medium.com/analytics-vidhya/illustrative-example-of-principal-component-analysis-pca-vs-linear-discriminant-analysis-lda-is-105c431e8907

Understanding stratified cross-validation. (n.d.). Cross Validated. https://stats.stackexchange.com/questions/49540/understanding-stratified-cross-validation

Vadapalli, P. (2022). Naive Bayes Explained: Function, Advantages & Disadvantages, Applications in 2023. upGrad Blog. https://www.upgrad.com/blog/naive-bayes-explained/#:~:text=Naive%20Bayes%20is%20suitable%20for,input%20variables%20than%20numerical%20variables.

Walk-Forward Matrix - StrategyQuant. (2021, August 10). StrategyQuant. https://strategyquant.com/doc/strategyquant/walk-forward-matrix/

What is Machine Learning? | IBM. (n.d.). https://www.ibm.com/topics/machine-learning#:~:text=the%20next%20step-,What%20is%20machine%20learning%3F,learn%2C%20gradually%20improving%20its%20accuracy

Wikipedia contributors. (2023). Cross-validation (statistics). Wikipedia. https://en.wikipedia.org/wiki/Cross-validation_%28statistics%29

Wikipedia contributors. (2023b). Walk forward optimization. Wikipedia. https://en.wikipedia.org/wiki/Walk_forward_optimization#:~:text=Walk%20Forward%20Analysis%20does%20optimization,Pardo

ChatGPT