

---

Professional Certificate in Artificial Intelligence for Welding Processes

## Unsupervised Learning Algorithms

---

Unsupervised learning refers to a family of algorithms that infer structure from data without the use of explicit target labels. In the context of welding processes, these techniques enable the discovery of patterns in sensor streams, high-dimensional images of weld pools, and spectroscopic signatures of metal vapour without requiring costly manual annotation. The terminology associated with unsupervised learning is extensive, and a solid grasp of each concept is essential for applying these methods to real-world welding applications.

**Dataset** – The collection of observations gathered from welding experiments. Each observation may correspond to a single weld pass, a time window of arc voltage, a spectrum, or a micro-graph image. The dataset is usually represented as a matrix  $X$  with  $n$  rows (observations) and  $p$  columns (features). In welding contexts,  $p$  can be very large when dealing with high-frequency voltage-current waveforms, multi-sensor arrays, or hyperspectral imaging.

**Feature** – An individual measurable property of a weld. Examples include peak current, average temperature, gas composition, pixel intensity in a weld-pool image, or the coefficient of a Fourier transform of the voltage signal. Features can be raw sensor outputs or engineered descriptors such as statistical moments, texture measures, or frequency-domain amplitudes.

**Observation** – A single instance in the dataset, often a weld trial or a time slice of a continuous welding process. An observation is described by a vector of feature values. In unsupervised learning, observations are the primary objects that the algorithm groups, separates, or projects.

**Similarity** – A quantitative assessment of how alike two observations are. Similarity is the inverse of a distance measure; the higher the similarity, the closer the observations are considered to be. In welding data, similarity may be defined based on Euclidean distance of voltage waveforms, cosine similarity of spectral vectors, or dynamic time-warping (DTW) distance for time-series signals.

**Distance metric** – A function that computes the dissimilarity between two observations. Common metrics include Euclidean distance, Manhattan distance, Mahalanobis distance, and cosine distance. Selecting an appropriate metric is critical for welding data because the physical meaning of the features (e.g., Temperature versus frequency) may require scaling or weighting to reflect domain knowledge.

**Feature scaling** – The process of adjusting feature ranges so that no single feature dominates the distance calculation. Two popular methods are standardization (subtracting the mean and dividing by the standard deviation) and min-max normalization (scaling values to the interval  $[0, 1]$ ). For welding signals, scaling may also involve whitening, where the covariance matrix of the data is transformed to the identity matrix, thereby decorrelating the features.

**Clustering** – The task of partitioning observations into groups, or clusters, such that observations within the

---

same cluster are more similar to each other than to those in other clusters. In welding, clustering can reveal distinct operating regimes (e.G., Stable arc, spatter-prone arc), identify groups of weld defects, or separate different material compositions based on sensor signatures.

**Centroid** – The arithmetic mean of all observations assigned to a cluster. In k-means clustering, each cluster is represented by its centroid, and the algorithm iteratively updates centroids and reassigns observations to the nearest centroid. For welding data, centroids may correspond to typical waveform shapes or average spectroscopic profiles for a specific defect type.

**k-means clustering** – A widely used partitioning algorithm that seeks to minimize the sum of squared distances between observations and their assigned centroids. The number of clusters  $k$  must be specified in advance. In welding applications,  $k$  might be chosen based on prior knowledge of the number of defect categories (e.G., Lack of fusion, porosity, crack) or determined empirically using validation techniques.

**Elbow method** – A heuristic for selecting  $k$  in k-means clustering. The method plots the total within-cluster sum of squares (WCSS) against increasing values of  $k$ . The point where the reduction in WCSS begins to plateau, resembling an “elbow,” suggests a suitable number of clusters. When applied to welding data, the elbow often corresponds to the point where adding another cluster provides little additional insight into defect differentiation.

**Silhouette score** – A measure of how well an observation fits within its cluster compared to the nearest neighboring cluster. The score ranges from  $-1$  to  $+1$ ; higher values indicate better clustering. In welding analysis, silhouette scores can be computed for each weld pass to assess the reliability of defect grouping, helping to flag ambiguous cases for further inspection.

**Hierarchical clustering** – An alternative to partitioning methods that builds a tree-like structure (dendrogram) representing nested groupings of observations. Two main strategies exist: Agglomerative (bottom-up) and divisive (top-down). In welding, hierarchical clustering is useful for exploring relationships among multiple defect types, as it reveals how fine-grained categories merge into broader groups.

**Linkage** – The rule used to compute the distance between clusters in hierarchical clustering. Common linkage criteria include single linkage (minimum distance), complete linkage (maximum distance), average linkage (average distance), and Ward’s method (increase in total variance). Ward’s method is often preferred for welding data because it tends to produce compact, spherical clusters that align with the physical intuition of defect categories.

**Dendrogram** – A graphical representation of the hierarchical clustering process, where the vertical axis shows the distance at which clusters merge. By cutting the dendrogram at a chosen height, a specific number of clusters can be extracted. In a welding quality-control setting, a dendrogram can illustrate how different weld defects are related, aiding in the development of targeted remediation strategies.

**Density-based clustering** – A class of algorithms that identify clusters as dense regions of points separated by areas of lower point density. This approach is particularly robust to noise and can discover clusters of arbitrary shape, which is advantageous for complex welding data where defect patterns may not be spherical.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – A seminal density-based algorithm that requires two parameters:  $\epsilon$  (the radius of a neighborhood) and  $\text{minPts}$  (the minimum number of points required to form a dense region). Points that belong to a dense region become core points; points reachable from core points become part of the same cluster, while points that cannot be reached are labeled as noise. In welding, DBSCAN can isolate outlier welds caused by transient disturbances, such as sudden voltage spikes, while grouping stable welds into coherent clusters.

OPTICS (Ordering Points To Identify the Clustering Structure) – An extension of DBSCAN that produces an augmented ordering of the data points, allowing the extraction of clusters at multiple density levels without committing to a single  $\epsilon$  value. For welding processes with varying signal intensities, OPTICS can reveal hierarchical density structures, helping to differentiate between subtle defect modes and severe anomalies.

Gaussian mixture model (GMM) – A probabilistic model that represents the data as a mixture of several Gaussian distributions, each corresponding to a cluster. The model parameters (means, covariances, and mixing coefficients) are estimated using the Expectation-Maximization (EM) algorithm. GMMs provide soft cluster assignments, giving the probability that a weld belongs to each defect class. This probabilistic output is valuable for risk-based decision making in welding automation.

Expectation-Maximization – An iterative optimization technique used to estimate the parameters of statistical models with latent variables. In the E-step, the algorithm computes the expected membership probabilities for each observation given the current model parameters; in the M-step, it updates the parameters to maximize the expected log-likelihood. For welding data, EM can handle missing sensor values by treating them as latent variables, improving the robustness of the clustering outcome.

Dimensionality reduction – The process of mapping high-dimensional data onto a lower-dimensional space while preserving essential structure. Reducing dimensionality helps visualize welding data, speeds up clustering, and mitigates the curse of dimensionality. Common techniques include Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP), and autoencoders.

Principal Component Analysis – A linear technique that identifies orthogonal directions (principal components) capturing the greatest variance in the data. The first principal component explains the largest amount of variance, the second explains the next largest, and so on. In welding, PCA can be used to compress thousands of voltage-current samples into a handful of components that still retain the dominant welding dynamics, facilitating downstream clustering or anomaly detection.

Eigenvector – A non-zero vector that, when multiplied by a square matrix, yields a scalar multiple of itself. In PCA, eigenvectors of the covariance matrix define the directions of the principal components. For welding data, the eigenvectors may reveal latent modes such as “arc stability” or “metal transfer frequency,” which are not directly observable but influence the measured signals.

Eigenvalue – The scalar factor associated with an eigenvector in the eigenvalue equation. In PCA, each eigenvalue quantifies the amount of variance captured by its corresponding eigenvector. Large eigenvalues indicate that the associated principal component carries significant information about the welding process;

---

small eigenvalues suggest noise or redundant dimensions.

**Loadings** – The coefficients that express each original feature as a linear combination of the principal components. Loadings help interpret the meaning of each component. For instance, a loading heavily weighted on high-frequency voltage components may indicate a component related to spatter generation in the welding arc.

**Variance explained** – The proportion of total data variance accounted for by a particular principal component or a set of components. In welding applications, a scree plot of variance explained can guide the selection of the number of components to retain, ensuring that essential process dynamics are preserved while discarding insignificant noise.

**t-Distributed Stochastic Neighbor Embedding (t-SNE)** – A non-linear dimensionality-reduction method that preserves local structure by converting pairwise similarities into probabilities and then minimizing the Kullback-Leibler divergence between the high-dimensional and low-dimensional distributions. T-SNE is especially effective for visualizing complex welding datasets where clusters may be intertwined. However, it is computationally intensive and sensitive to hyperparameters such as perplexity and learning rate.

**Uniform Manifold Approximation and Projection (UMAP)** – A newer non-linear technique that approximates the manifold structure of the data and produces a low-dimensional embedding. UMAP often yields clearer global structure than t-SNE and runs faster on large welding datasets. It can be used to visualize the distribution of weld defects across multiple sensor modalities, facilitating the identification of outlier welds.

**Autoencoder** – A type of neural network that learns to reconstruct its input after passing through a bottleneck layer of reduced dimensionality. The encoder part compresses the input into a latent representation; the decoder reconstructs the original input from this latent code. In welding, autoencoders can learn compact representations of high-frequency voltage waveforms or multi-spectral images, which can then be clustered or examined for anomalies.

**Latent space** – The low-dimensional representation learned by an autoencoder or other generative model. Points in the latent space capture the essential characteristics of the original welding data. Similar latent vectors correspond to similar weld signatures, enabling clustering directly in the latent space and often improving separation of defect categories.

**Reconstruction error** – The difference between the original input and its reconstruction produced by an autoencoder, typically measured by mean-squared error (MSE) or mean absolute error (MAE). High reconstruction error can signal an anomalous weld that deviates from the patterns learned during training, making it a useful metric for unsupervised anomaly detection in welding processes.

**Anomaly detection** – The identification of observations that differ significantly from the majority of the data. In welding, anomalies may correspond to defective welds, sensor malfunctions, or unexpected process disturbances. Unsupervised methods such as one-class SVM, isolation forest, or autoencoder-based reconstruction error are commonly employed to flag such cases without requiring labeled defect data.

**Outlier** – An observation that lies far from the bulk of the data distribution. Outliers can be genuine defects

---

(e.G., A weld with an unexpected crack) or they may be spurious measurements caused by noise or sensor drift. Distinguishing between these two types often requires domain expertise and may involve additional post-processing steps.

**Novelty detection** – Similar to anomaly detection, but focuses on identifying new, previously unseen patterns rather than isolated outliers. In welding, novelty detection can be used to discover emerging defect modes as new materials or welding parameters are introduced, enabling proactive adjustments to process controls.

**Manifold learning** – A family of techniques that assume data lie on a low-dimensional manifold embedded in a high-dimensional space. Methods such as Isomap, Locally Linear Embedding (LLE), and spectral clustering exploit this assumption to uncover the intrinsic geometry of welding data. Manifold learning can reveal smooth trajectories of welding parameters, aiding in the design of optimal process windows.

**Spectral clustering** – An algorithm that constructs a similarity graph from the data, computes the graph Laplacian, and then performs eigen-decomposition to obtain a low-dimensional embedding. K-means is then applied to the embedding to produce clusters. Spectral clustering excels at detecting non-convex cluster shapes, which is advantageous for welding datasets where defect clusters may be elongated or intertwined due to overlapping process signatures.

**Graph Laplacian** – A matrix representation of a graph that captures the connectivity of nodes (observations) based on similarity. The normalized Laplacian is often used in spectral clustering to ensure stability and scale invariance. In welding, constructing a similarity graph from multi-sensor data and applying the Laplacian can highlight groups of welds that share common physical characteristics.

**Community detection** – A concept from network analysis where the goal is to find groups of nodes that are more densely connected internally than with the rest of the graph. Algorithms such as modularity maximization or the Louvain method can be applied to the similarity graph of welding observations, yielding clusters that reflect underlying process regimes.

**Clustering validation** – The set of techniques used to assess the quality and reliability of a clustering result. Validation can be internal (based solely on the data) or external (using external labels, if available). For welding, internal validation is often necessary because ground-truth defect labels may be scarce or expensive to obtain.

**Internal validation** – Measures that evaluate clustering quality without reference to external information. Common internal metrics include silhouette score, Davies-Bouldin index, Calinski-Harabasz index, and within-cluster sum of squares. In welding, internal validation helps determine whether a clustering solution meaningfully separates different weld signatures.

**External validation** – Metrics that compare the clustering assignments to known class labels. Examples are Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Fowlkes-Mallows index. When a limited set of labeled weld defects is available, external validation can quantify how well the unsupervised clusters align with expert classifications.

**Adjusted Rand Index** – A statistic that measures the similarity between two data clusterings, corrected for chance. ARI values range from  $-1$  (no agreement) to  $1$  (perfect agreement). In welding research, ARI can be used to benchmark a new clustering algorithm against a baseline that uses manually labeled defect categories.

**Mutual Information** – A measure of the amount of information shared between two random variables, such as the true defect labels and the cluster assignments. Normalized Mutual Information scales the value to the interval  $[0, 1]$ , where  $1$  indicates perfect correlation. MI is useful for evaluating the correspondence between unsupervised clusters and known welding defect types.

**Scalability** – The ability of an algorithm to handle increasing data size without excessive computational or memory demands. Many welding monitoring systems generate streams of data at kilohertz rates, leading to massive datasets. Algorithms like Mini-Batch k-means, approximate DBSCAN, and stochastic gradient descent-based autoencoders are designed for scalability, enabling real-time analysis of welding processes.

**Hyperparameter tuning** – The process of selecting optimal values for algorithm parameters (e.g.,  $k$  in k-means,  $\epsilon$  in DBSCAN, number of layers in an autoencoder). In welding, hyperparameter selection often involves domain-specific constraints such as allowable processing time, sensor resolution, and acceptable false-positive rates for defect detection. Grid search, random search, and Bayesian optimization are common strategies for systematic tuning.

**Noise sensitivity** – The degree to which an algorithm's output is affected by random fluctuations or measurement errors. Some clustering methods, like k-means, are highly sensitive to noise because outliers can distort centroid positions. Density-based methods (DBSCAN, OPTICS) and robust statistical models (Gaussian mixture models with full covariance) are more tolerant of noisy welding data.

**Interpretability** – The extent to which the results of an algorithm can be understood and acted upon by practitioners. In welding, interpretability is crucial because engineers need to translate cluster assignments into actionable process adjustments. Linear methods (PCA, k-means) are generally more interpretable than deep neural network embeddings, but the latter may uncover subtle patterns that are otherwise hidden.

**Feature engineering** – The practice of creating informative features from raw sensor data. For welding, this may involve extracting statistical moments (mean, variance, skewness), frequency-domain features (FFT amplitudes, spectral centroids), wavelet coefficients, or texture descriptors from weld-pool images. Thoughtful feature engineering can greatly improve the performance of unsupervised algorithms by emphasizing the most relevant physical phenomena.

**Time-series segmentation** – Dividing a continuous welding signal into meaningful segments, often based on changes in statistical properties or detected events (e.g., Electrode contact, arc ignition). Segmentation enables the application of clustering to each segment separately, allowing the discovery of temporal patterns such as intermittent spatter or periodic voltage fluctuations.

**Dynamic time-warping (DTW)** – A distance measure specialized for time-series data that aligns sequences by allowing non-linear stretching of the time axis. DTW is particularly useful for welding voltage or current waveforms where the same physical event may occur at slightly different times across welds. Clustering with

---

DTW can group welds exhibiting similar temporal dynamics even when their raw signals are misaligned.

Kernel methods – Techniques that implicitly map data into a higher-dimensional feature space using a kernel function, enabling linear algorithms to capture non-linear relationships. Kernel k-means and kernel PCA are examples. In welding, kernels such as the Gaussian (RBF) kernel can capture complex interactions between temperature and gas flow, improving clustering quality without explicit feature transformations.

Self-Organizing Map (SOM) – A type of neural network that projects high-dimensional data onto a two-dimensional grid while preserving topological relationships. SOMs provide a visual map where similar weld signatures occupy neighboring cells. This can be used for exploratory analysis of welding datasets, revealing clusters of similar defect types or process conditions.

Isolation Forest – An ensemble method for anomaly detection that isolates observations by randomly partitioning the data. Anomalies require fewer splits to isolate, leading to lower average path lengths. Isolation Forests are computationally efficient and work well with high-dimensional welding data, making them suitable for online monitoring of arc stability.

One-Class Support Vector Machine (One-Class SVM) – A model that learns a decision boundary that encompasses the majority of the data, treating points outside the boundary as anomalies. The method is based on kernel functions and can capture complex shapes. In welding, One-Class SVM can be trained on normal weld signatures and later used to flag abnormal welds that may indicate defects.

Batch processing – The practice of collecting a set of weld observations and processing them together. Batch processing allows the use of algorithms that require the entire dataset in memory, such as full-matrix PCA or hierarchical clustering. However, in real-time welding control, batch processing may introduce latency, prompting the need for incremental or online methods.

Incremental learning – Algorithms that update their model as new data arrive, without retraining from scratch. Incremental PCA, online k-means, and streaming autoencoders are examples. For welding, incremental learning enables continuous adaptation to changing conditions (e.g., Wear of electrodes, variations in shielding gas) while maintaining up-to-date clustering models.

Streaming data – A continuous flow of data generated by sensors attached to welding equipment. Streaming data requires algorithms that can operate under limited memory and computational constraints. Techniques such as reservoir sampling, sliding windows, and online clustering are essential for handling streaming welding data.

Visualization – The practice of representing high-dimensional data in a form that can be intuitively understood. Common visualizations for unsupervised learning include scatter plots of the first two principal components, t-SNE or UMAP embeddings, dendrograms for hierarchical clustering, and heatmaps of similarity matrices. In welding, visualizations help engineers quickly spot abnormal patterns, assess the separation between defect clusters, and communicate findings to stakeholders.

Heatmap – A graphical representation of a matrix where each cell's color indicates the magnitude of a value, such as a similarity or distance measure. Heatmaps of the pairwise distances between weld signatures

---

can reveal natural groupings and highlight outliers.

Scatter plot matrix – A grid of scatter plots showing pairwise relationships between features. For welding data, a scatter plot matrix can expose correlations between voltage peak, current ripple, and gas flow rate, informing the selection of distance metrics and clustering algorithms.

Cluster assignment – The label indicating which cluster an observation belongs to. In welding, cluster assignments can be mapped back to the physical process to identify which welds belong to a “stable arc” cluster versus a “spatter-prone” cluster, supporting targeted process adjustments.

Soft clustering – An approach where observations have a probability of belonging to each cluster rather than a hard assignment. Gaussian mixture models provide soft clustering, allowing a weld to be partially associated with multiple defect categories. Soft assignments can be used to compute confidence scores for automated inspection systems.

Hard clustering – Traditional clustering where each observation is assigned to a single cluster. Methods such as k-means and DBSCAN produce hard assignments, which are often easier to interpret in welding quality-control contexts.

Cluster centroid drift – The phenomenon where centroids shift over time as the underlying data distribution changes. In continuous welding operations, drift may indicate gradual changes in electrode wear or shielding gas composition. Monitoring centroid drift can serve as an early warning system for process degradation.

Feature importance – Quantitative measures indicating how much each feature contributes to the formation of clusters. Techniques such as permutation importance, SHAP values, or inspecting PCA loadings provide insight into which sensor measurements drive the clustering. In welding, identifying important features helps prioritize sensor placement and data acquisition.

Permutation importance – A method that evaluates the decrease in clustering performance when a feature’s values are randomly shuffled. A large drop indicates that the feature is critical for distinguishing clusters. Applied to welding data, permutation importance can reveal that, for example, the high-frequency component of the current waveform is a key discriminator between defect types.

Silhouette plot – A visual tool that displays the silhouette coefficient for each observation, grouped by cluster. The plot shows the distribution of silhouette values, helping to assess cluster cohesion and separation. In welding, a silhouette plot can quickly indicate whether a particular defect cluster is well-defined or if it contains ambiguous welds.

Cluster stability – The degree to which a clustering solution remains consistent under variations in the data (e.G., Subsampling, noise injection) or algorithmic parameters. Stability analysis can be performed using bootstrap resampling or by comparing clusterings obtained with different random seeds. For welding, stable clusters suggest robust groupings of defect patterns that are reliable for decision making.

Bootstrap resampling – A statistical technique that creates many synthetic datasets by sampling with

---

replacement from the original data. Clustering is performed on each bootstrap sample, and the consistency of cluster assignments is measured. High consistency indicates that the identified weld clusters are not artifacts of a particular data split.

Cross-validation – A method for assessing model performance by partitioning data into training and validation subsets. While traditionally associated with supervised learning, cross-validation can be adapted for unsupervised tasks by evaluating internal validation metrics on held-out data. In welding, cross-validation helps ensure that the clustering model generalizes to unseen welds.

Model selection – The process of choosing the most appropriate algorithm and its hyperparameters based on validation results. For welding data, model selection may involve comparing k-means, hierarchical clustering, DBSCAN, and Gaussian mixture models, each evaluated with silhouette scores, ARI (when labels are available), and computational cost.

Computational complexity – An estimate of the time and memory resources required by an algorithm as a function of the number of observations  $n$  and the number of features  $p$ . K-means typically scales as  $O(np^2)$ , while hierarchical clustering scales as  $O(n^2)$  in time and  $O(n^2)$  in memory, making the latter impractical for very large welding datasets without approximation techniques.

Approximation techniques – Methods that reduce computational load by sacrificing exactness for speed. Examples include using the Nyström method for approximating the eigen-vectors in spectral clustering, employing Mini-Batch k-means instead of full-batch k-means, or using fast approximate nearest neighbor (ANN) search for density-based clustering. In welding, these techniques enable real-time analysis of high-frequency sensor streams.

Fast approximate nearest neighbor – Algorithms such as Annoy, HNSW, or FLANN that quickly retrieve approximate nearest neighbors in high-dimensional spaces. These are essential for scaling DBSCAN or OPTICS to large welding datasets, where exact neighbor searches would be prohibitive.

Graph-based clustering – A family of methods that construct a graph from the data and then partition the graph into communities. Spectral clustering, community detection, and Markov clustering fall under this umbrella. Graph-based approaches are powerful for welding data because they naturally incorporate similarity information derived from multi-sensor fusion.

Multi-sensor fusion – The integration of data from different sensor modalities (e.g., Voltage, current, acoustic emission, infrared imaging) into a unified representation. Fusion can be performed at the feature level (concatenating feature vectors), at the decision level (combining clustering results), or using deep learning models that learn joint embeddings. Effective fusion often improves clustering separation of weld defects.

Feature concatenation – A straightforward fusion technique where features from each sensor are combined into a single high-dimensional vector. While simple, concatenation can lead to the curse of dimensionality; therefore, dimensionality-reduction methods such as PCA or autoencoders are frequently applied afterward.

Decision-level fusion – A strategy where clustering is performed separately on each sensor modality, and

the resulting cluster assignments are combined using voting or weighted averaging. Decision-level fusion can exploit the strengths of each sensor while mitigating individual weaknesses, leading to more robust defect detection.

**Joint embedding** – A learned representation where data from multiple modalities are projected into a common latent space. Techniques such as multimodal autoencoders, canonical correlation analysis (CCA), or contrastive learning can produce joint embeddings that capture the shared structure of welding signals and images. Clustering in the joint space often yields more coherent groups than clustering each modality alone.

**Canonical correlation analysis** – A statistical method that finds linear combinations of two sets of variables (e.g., Voltage features and spectral features) that are maximally correlated. CCA can be used to align the two modalities before clustering, ensuring that the resulting clusters reflect joint behavior.

**Contrastive learning** – A self-supervised approach where a model learns to bring similar pairs of data points closer together in the embedding space while pushing dissimilar pairs apart. For welding, similar pairs could be different views of the same weld (e.g., Voltage waveform and infrared image), and dissimilar pairs could be unrelated welds. The learned embeddings can then be clustered.

**Data augmentation** – The creation of synthetic training examples by applying transformations to existing data (e.g., Adding noise, scaling, time-shifting). Augmentation improves the robustness of unsupervised models, especially autoencoders, by exposing them to a wider variety of realistic variations in welding signals.

**Regularization** – Techniques that constrain model complexity to prevent overfitting. In the context of autoencoders for welding data, regularization may involve adding an L2 penalty on weights, applying dropout, or enforcing sparsity in the latent representation. Regularization helps ensure that the learned features capture genuine process characteristics rather than sensor noise.

**Dropout** – A stochastic regularization method where a random subset of neurons is deactivated during each training iteration. Dropout forces the network to develop redundant representations, which can improve the generalization of autoencoders used for welding data compression.

**Batch normalization** – A technique that normalizes the activations of each layer to have zero mean and unit variance, improving training stability. When training deep autoencoders on large welding datasets, batch normalization can accelerate convergence and reduce sensitivity to the choice of learning rate.

**Learning rate** – A hyperparameter that controls the size of weight updates during gradient-based optimization. Choosing an appropriate learning rate is crucial for training autoencoders on welding data; too large a learning rate may cause divergence, while too small a learning rate can lead to excessively long training times.

**Optimizer** – An algorithm that updates model parameters based on gradients. Common optimizers include stochastic gradient descent (SGD), Adam, and RMSprop. Adam is often favored for training autoencoders on welding data because it adapts learning rates for each parameter, handling the diverse scales of sensor

---

features.

**Early stopping** – A regularization strategy that halts training when performance on a validation set stops improving. Early stopping prevents overfitting of autoencoders to the training welding data, preserving the model's ability to generalize to new welds.

**Reconstruction loss** – The objective function minimized during autoencoder training, typically the mean squared error between input and reconstructed output. Minimizing reconstruction loss forces the encoder to capture the most salient aspects of welding signals while discarding noise.

**Latent dimensionality** – The size of the bottleneck layer in an autoencoder. Selecting an appropriate latent dimensionality balances compression against information loss. For welding data, a latent space of 10–30 dimensions often suffices to capture the essential dynamics of voltage-current waveforms while reducing computational burden for downstream clustering.

**Clustering on latent space** – Performing clustering directly on the low-dimensional representations learned by an autoencoder. This approach leverages the autoencoder's ability to denoise and compress the data, often resulting in clearer separation of weld defect clusters than clustering on raw high-dimensional features.

**Out-of-sample projection** – The ability to map new, unseen observations into the learned latent space without retraining the model. In welding, this property enables a trained autoencoder to embed new weld signatures in real time, after which the clustering model can assign them to the appropriate defect group.

**Transfer learning** – Reusing a model trained on one dataset for a different but related dataset. In welding, a model trained on a particular alloy and welding configuration can be fine-tuned on a new alloy, accelerating the development of unsupervised analysis pipelines for novel materials.

**Domain adaptation** – Techniques that adjust models to account for distribution shifts between source and target domains. For welding, domain adaptation may be required when moving from laboratory-controlled welding experiments to field-deployed robotic welding lines, where sensor noise characteristics and operating conditions differ.

**Semi-supervised learning** – A hybrid approach that leverages a small amount of labeled data together with a large pool of unlabeled data. In welding, a few expertly annotated defect samples can guide clustering, for example by seeding the initial centroids in k-means or by constraining the Gaussian mixture model with partial label information.

**Constrained clustering** – Clustering methods that incorporate prior knowledge in the form of must-link or cannot-link constraints. Must-link constraints force two observations to be placed in the same cluster, while cannot-link constraints prevent them from sharing a cluster. In welding, constraints can be derived from expert knowledge (e.g., Two welds known to share the same defect type must be linked) to improve clustering fidelity.

**Active learning** – An interactive process where the algorithm selects the most informative observations for

---

labeling by an expert. Active learning can be applied to unsupervised welding analysis by asking a technician to label a few ambiguous welds, thereby refining the clustering model with minimal labeling effort.

**Evaluation pipeline** – The sequence of steps used to assess unsupervised models, typically including data preprocessing, feature extraction, dimensionality reduction, clustering, validation, and visualization. A well-designed pipeline ensures reproducibility and facilitates comparison of different algorithms on welding datasets.

**Reproducibility** – The ability to obtain the same results when the analysis is repeated under identical conditions. In welding research, reproducibility is achieved by fixing random seeds, documenting preprocessing steps, and preserving versioned code and data.

**Random seed** – An initial value that determines the sequence of pseudo-random numbers generated by an algorithm. Setting a random seed ensures that stochastic processes (e.g., Centroid initialization in k-means) produce consistent results across runs, aiding reproducibility.

**Version control** – Systems such as Git that track changes to code and configuration files. Maintaining version control for welding data analysis scripts enables collaborative development and facilitates rollback to previous experiment configurations.

**Data provenance** – Metadata that records the origin, collection method, and processing history of each dataset. In welding, provenance information includes sensor calibration dates, welding parameters (current, voltage, travel speed), and environmental conditions, all of which are essential for interpreting unsupervised learning outcomes.

**Scikit-learn** – A popular Python library that provides implementations of many unsupervised algorithms (k-means, DBSCAN, PCA, t-SNE, etc.). Scikit-learn's consistent API simplifies experimentation on welding datasets, allowing rapid prototyping and benchmarking of different clustering approaches.

**TensorFlow and PyTorch** – Deep learning frameworks used to build autoencoders, variational autoencoders, and contrastive learning models for welding data. Both libraries support GPU acceleration, which is valuable for training large models on high-frequency sensor streams.

**GPU acceleration** – The use of graphics processing units to speed up computationally intensive tasks such as matrix multiplications in deep learning.