NSCT AI, Machine Learning & Data Analytics Cheat Sheet

By the NSCT Exam Editorial Team · Updated 2026-11-24 · Facts checked against official sources on 2026-11-24

This page puts the metrics, formulas and comparisons behind most AI, Machine Learning & Data Analytics questions on one sheet. The area carries 10% of the NSCT per HEC's Areas of Competencies, and 2026 candidates averaged 50.8% accuracy in it per the HEC NSCT report and APP. Print it (or save it as a PDF from your browser) and keep it next to your practice.

Key definitions

Term	One-line meaning
AI ⊃ ML ⊃ DL	Deep learning is a subset of machine learning, which is a subset of AI
Supervised	Learns from labelled data (regression = continuous target, classification = discrete)
Unsupervised	No labels: clustering, dimensionality reduction, association rules
Reinforcement	Agent learns from rewards
Self-supervised	Labels generated from the data itself (masked words, next token)
Validation set	Tunes hyperparameters; the test set is used once at the end
Data leakage	Information from outside the training data reaches the model
Overfitting	Low training error, high test error (high variance)
Underfitting	High error on both (high bias)
Hyperparameter	Set before training (k, learning rate, depth), not learned
Epoch	One full pass over the training data
Data drift / concept drift	P(X) changes / P(y|X) changes

Analytics types

Type	Question
Descriptive	What happened?
Diagnostic	Why did it happen?
Predictive	What will happen?
Prescriptive	What should we do?

Core formulas

Formula	Expression	Worked check
Bayes	$P(A\|B) = P(B\|A)\,P(A) \div P(B)$	
Matrix product shape	$(m \times n)(n \times p) = (m \times p)$	
2 × 2 determinant	$ad - bc$	det 0 → no inverse
Gradient descent	$w \leftarrow w - \eta \times \partial L/\partial w$	η = learning rate
Min-max scaling	$(x - \min) \div (\max - \min)$	x 30, range 10–50 → 0.5
Standardisation (z)	$(x - \text{mean}) \div \text{sd}$	x 70, mean 60, sd 5 → 2
Robust scaling	$(x - \text{median}) \div \text{IQR}$	Least outlier-sensitive
Euclidean distance	$\sqrt{\Sigma(x_i - y_i)^2}$	(3, 4) to (6, 8) = 5
Manhattan distance	$\Sigma\|x_i - y_i\|$	(3, 4) to (6, 8) = 7
Entropy	$-\Sigma\, p_i \log_2 p_i$	50/50 node = 1

Formula	Expression	Worked check
Gini	$1 - \Sigma\, p_i^2$	50/50 node = 0.5
Information gain	Parent entropy − weighted child entropy	
SVM margin	$2 \div \|\|w\|\|$	Larger C → narrower margin
AdaBoost weight	$\alpha = \frac{1}{2} \ln((1 - e) \div e)$	$e = 0.2 \rightarrow \alpha \approx 0.693$
TF-IDF	$tf \times \log(N \div df)$	Term in every document → idf 0
PCA explained variance	eigenvalue ÷ sum of eigenvalues	
Bootstrap coverage	$1 - 1/e \approx 63.2\%$ distinct rows	36.8% out-of-bag

Confusion-matrix metrics

Worked on TP 40, FP 10, FN 20, TN 130 (200 rows).

Metric	Formula	Value
Accuracy	(TP + TN) ÷ total	0.85
Precision	TP ÷ (TP + FP)	0.80
Recall (sensitivity, TPR)	TP ÷ (TP + FN)	0.667
Specificity (TNR)	TN ÷ (TN + FP)	0.929
False positive rate	FP ÷ (FP + TN)	0.071
F1	2PR ÷ (P + R)	0.727

AUC = probability that a random positive is ranked above a random negative (0.5 = random guessing).

Neural network numbers

Item	Rule	Worked check
Conv output size	$\lfloor (n + 2p - f) \div s \rfloor + 1$	32, f 5, p 0, s 1 → 28; p 2 → 32
Pooling output	Same formula, no weights, depth unchanged	28, 2 × 2, s 2 → 14
Conv output depth	Number of filters	
Conv layer parameters	(f × f × channels + 1) × filters	3 × 3 × 3, 16 filters → 448
Dense parameters	inputs × outputs + outputs (biases)	784-128-10 → 101,770
Iterations per epoch	samples ÷ batch size	10,000 ÷ 100 = 100
Attention	$\text{softmax}(QK^T \div \sqrt{d_k})\, V$	$O(n^2)$ in sequence length

Activation	Range	Typical use
Sigmoid	(0, 1); derivative peaks at 0.25	Binary output
Tanh	(−1, 1)	Zero-centred hidden layers
ReLU	[0, ∞)	Default hidden layer
Softmax	Outputs sum to 1	Single-label multi-class output
Linear	Any real	Regression output

Task	Output layer	Loss
Regression	Linear	MSE or MAE
Binary classification	Sigmoid	Binary cross-entropy
Multi-class, one label	Softmax	Categorical cross-entropy

Statistics for EDA

Fact	Value
Right (positive) skew	Mean > median (income)
Left skew	Mean < median
Empirical rule	~68%, 95%, 99.7% within 1, 2, 3 sd
Sample vs population sd	Divide by n − 1 vs n
Pearson vs Spearman	Linear vs any monotonic relationship

Big data and MLOps

Item	Fact
5 Vs	Volume, velocity, variety, veracity (quality), value
HDFS block	128 MB default (Hadoop 2.x, 3.x), replication 3
NameNode / DataNode	Metadata / blocks; Secondary NameNode is not a failover
MapReduce	Map → shuffle and sort → reduce
Spark	Transformations lazy; actions (count, collect) execute
Canary vs shadow	Some users see the new model vs none do (outputs only logged)

X vs Y quick distinctions

Pair	Difference
Precision vs recall	Divide by predicted positives vs actual positives
Bagging vs boosting	Parallel, cuts variance vs sequential, cuts bias
K-means vs k-NN	Unsupervised clustering vs supervised classifier
L1 (Lasso) vs L2 (Ridge)	Can zero out weights (selection) vs shrinks weights
Selection vs extraction	Keeps original columns vs builds new ones (PCA)
Filter vs wrapper vs embedded	Statistic only vs retrains the model vs built into training
Stemming vs lemmatisation	Chops suffixes (studies → studi) vs dictionary form (better → good)
Skip-gram vs CBOW	Centre predicts context vs context predicts centre
LSTM vs GRU	Forget, input, output gates vs reset, update gates
BERT vs GPT	Encoder, masked LM vs decoder, next-token
Semantic vs instance segmentation	Class per pixel vs mask per object
One-stage vs two-stage detector	YOLO, SSD: faster vs Faster R-CNN: usually more accurate
Single vs complete linkage	Minimum distance (chaining) vs maximum (compact)
Equal opportunity vs equalised odds	Equal TPR vs equal TPR and FPR

Classic traps

1. Swapping the precision and recall denominators.
2. Trusting accuracy on imbalanced data; 95% negatives give 95% for "always negative".
3. Scaling, oversampling or SMOTE before the train-test split (leakage).
4. Saying bagging reduces bias; it mainly reduces variance.
5. Picking k in k-means at minimum WCSS; WCSS always falls as k rises, so use the elbow or silhouette.
6. Forgetting biases when counting parameters.
7. Forgetting padding or the floor in the conv output formula.
8. Reading r = 0 as "no relationship"; $y = x^2$ can give r = 0.

9. Calling logistic regression a regression model; it is a classifier.
10. Thinking bag-of-words handles negation; "not good" still contains "good".
11. Calling wrong or duplicate values "variety"; that is veracity.
12. Thinking removing names anonymises data; quasi-identifiers can re-identify people.

Keep going

- Full notes with worked MCQs: AI & ML revision notes
- All 17 topics: AI & ML MCQ hub
- Deep practice: confusion matrix, overfitting and underfitting, model evaluation and deep learning

Next step

Check your recall in a timed subject-wise mock test, then take a full NSCT mock test.