

■ Data Science, Machine Learning & AI Study Plan (20 Weeks)

This plan is designed for a Computer Science student with a strong base in programming, data structures, algorithms, and some familiarity with SQL/DBMS. You will build solid foundations in Data Science, Machine Learning, and AI concepts in 20 weeks, studying ~4 hours/day for 5 days a week. You'll practice in Google Colab, but implement final projects locally.

Weeks 1–2: Foundations & Data Science Basics

- Revise SQL & DBMS (Normalization, Indexing, Joins).
- Book: 'Learning SQL' – key chapters.
- Python refresher: Numpy, Pandas, Matplotlib (Wes McKinney, Python for Data Analysis).
- Statistics: Descriptive stats, probability basics, hypothesis testing.
- Book: Practical Statistics for Data Scientists – Chapters 1–3.

Weeks 3–4: Regression Models

- Linear Regression & Logistic Regression – concepts + math.
- Book: ISL (Ch. 3 & 4).
- PRML reference: Ch. 1–4 for linear models & Bayesian view.
- Practice: implement regression models in scikit-learn.
- Mini-project: Predict student exam scores / house prices.

Weeks 5–6: Classification & Evaluation

- Decision Trees, Random Forest, SVM basics.
- Model evaluation: cross-validation, bias-variance tradeoff, ROC/AUC.
- ISL Ch. 5–9.
- PRML: SVMs & Bayesian classification.
- Practice Kaggle classification problem (Titanic).

Weeks 7–8: Feature Engineering & Unsupervised Learning

- Dimensionality Reduction: PCA, LDA.
- Clustering: K-means, Hierarchical.
- Feature engineering, normalization, encoding.
- ISL Ch. 10, 12.

- PRML: PCA + Latent Variable Models.
- Project: customer segmentation with clustering.

Weeks 9–10: Advanced ML

- Ensemble methods: Boosting, Bagging, Gradient Boosting.
- Model interpretability: SHAP, LIME.
- Hyperparameter tuning (GridSearch, RandomSearch).
- Hands-On ML (Géron) – relevant chapters.
- Mini-project: build an ensemble model for Kaggle dataset.

Weeks 11–12: Neural Networks Basics

- Intro to Neural Nets – perceptrons, backpropagation, gradient descent.
- Goodfellow DL Book Ch. 1–6.
- PRML Neural Networks section.
- Implement basic NN in Keras/TensorFlow (MNIST).

Weeks 13–14: Deep Learning Applications

- CNNs for images, RNNs for sequences.
- Hands-On ML – CNN & RNN chapters.
- Goodfellow Ch. 9–10.
- Project: Image classifier (CIFAR-10) or sentiment analysis.

Weeks 15–16: Data Science Integration

- Data storytelling & visualization.
- Book: Storytelling with Data (skim).
- Project: End-to-end ML pipeline with EDA → model → viz.
- SQL integration + Pandas pipelines.

Weeks 17–18: AI & Advanced Topics

- Intro to Reinforcement Learning (Goodfellow Ch. 16 overview).
- Ethics, fairness, bias in AI.
- Exploring transformers basics (BERT/GPT concepts).
- Mini-project: simple RL or NLP task.

Weeks 19–20: Capstone Projects

- Capstone 1: DS Project (EDA-heavy, predictive modeling).
- Capstone 2: ML/DL Project (image or text data).
- Document and push to GitHub.
- Prepare portfolio for internships/jobs.

■■ Notes: - Use Google Colab for experimenting, but finalize projects locally for performance + GitHub documentation. - PRML should be treated as a reference, not a linear read-through. - Balance reading with practice: every theory concept → implement a small example in Colab. - Aim for 2–3 strong projects by the end.