Machine Learning Mastery: 6-Month Study Plan

For: Mads
Goal: Transition from Data Analyst to Data Scientist / ML Engineer
Timeline: December 2024 → June 2025
Weekly commitment: 20-35 hours (3-6 hours weekday evenings + weekends)

Your Starting Point (Benchmark Summary)

Area	Current Level	Target Level
Python fundamentals	⭐⭐⭐	⭐⭐⭐⭐
Data wrangling (pandas/SQL)	⭐⭐⭐⭐	⭐⭐⭐⭐ (maintain)
Statistics & probability	⭐⭐	⭐⭐⭐⭐
Linear algebra	⭐	⭐⭐⭐
ML theory & intuition	⭐⭐	⭐⭐⭐⭐
Practical sklearn	⭐⭐	⭐⭐⭐⭐
Deep learning basics	⭐	⭐⭐⭐
MLOps / Production	⭐⭐	⭐⭐⭐

Resources Overview

Primary Video Courses

Resource	Cost	Use For
StatQuest (YouTube)	Free	Statistics & ML intuition - Josh Starmer explains everything brilliantly
3Blue1Brown (YouTube)	Free	Linear algebra & calculus visualization
Andrew Ng's ML Specialization (Coursera)	~€40/month or audit free	Foundational ML theory
fast.ai Practical Deep Learning	Free	Hands-on deep learning

Resource	Cost	Use For
Kaggle Learn	Free	Quick practical mini-courses

Books

For bedtime reading (conceptual):

- *Naked Statistics* by Charles Wheelan - Accessible, entertaining stats foundation
- *The Hundred-Page Machine Learning Book* by Andriy Burkov - Concise ML overview
- *Approaching (Almost) Any Machine Learning Problem* by Abhishek Thakur - Practical patterns

For deep study (technical):

- *An Introduction to Statistical Learning (ISLR)* - Free PDF at statlearning.com - THE stats/ML textbook
- *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow* by Aurélien Géron - Best practical ML book

Practice Platforms

Platform	Cost	Purpose
Kaggle	Free	Competitions, datasets, notebooks
LeetCode	Free/Premium	Coding interviews (you know this)
StrataScratch	Free tier	DS interview questions
MLOps Community	Free	Stay current on production ML

The 6-Month Curriculum

Phase 1: Foundations (Weeks 1-8)

The goal here is to rebuild your mathematical and statistical foundation while refreshing sklearn patterns. This is the "eat your vegetables" phase - not glamorous, but essential.

Month 1: Statistics & Probability Reset

Week 1: Descriptive Statistics & Distributions

Video (1.5-2 hrs/day):

- StatQuest: Histograms, Mean/Median/Mode, Standard Deviation
- StatQuest: The Normal Distribution
- StatQuest: Sampling from a Distribution
- Khan Academy: Probability basics (if needed)

Reading (bedtime):

- Start *Naked Statistics* chapters 1-4

Hands-on (weekends):

- Create a Jupyter notebook exploring distributions
- Generate random samples from normal, uniform, exponential distributions
- Visualize with histograms, calculate descriptive stats manually vs. with numpy/scipy
- Use your transaction data intuition: "What distribution do transaction amounts follow?"

Milestone: Can explain variance vs. standard deviation vs. standard error without looking it up.

Week 2: Probability Fundamentals

Video:

- StatQuest: Probability vs. Likelihood
- StatQuest: Bayes' Theorem
- 3Blue1Brown: Bayes theorem (visual intuition)

Reading:

- *Naked Statistics* chapters 5-7
- ISLR Chapter 2 (Statistical Learning overview)

Hands-on:

- Implement Bayes' theorem from scratch for a simple spam classifier (just the math, no sklearn)
- Calculate conditional probabilities on a real dataset

Milestone: Can explain what P(A|B) means and compute it manually.

Week 3: Hypothesis Testing & P-values

Video:

- StatQuest: Hypothesis Testing and p-values (WATCH THIS MULTIPLE TIMES)
- StatQuest: P-values: What they are and how to interpret them
- StatQuest: Statistical Power
- StatQuest: Type I and Type II errors

Reading:

- *Naked Statistics* chapters 8-10
- ISLR sections on inference

Hands-on:

- Perform t-tests, chi-squared tests on real data
- Deliberately create scenarios where you reject/fail to reject null hypothesis
- Visualize sampling distributions

Milestone: Can explain p-value correctly (not "probability the hypothesis is true").

Week 4: Confidence Intervals & Statistical Inference

Video:

- StatQuest: Confidence Intervals
- StatQuest: Bootstrapping
- StatQuest: Maximum Likelihood

Reading:

- ISLR Chapter 2 continued
- Finish *Naked Statistics*

Hands-on:

- Implement bootstrap confidence intervals from scratch

- Compare to scipy.stats confidence intervals

Project checkpoint: Create a "Statistics Fundamentals" notebook that you could show an interviewer, demonstrating understanding of distributions, hypothesis testing, and confidence intervals with real data examples.

Month 2: Linear Algebra + sklearn Patterns

Week 5: Linear Algebra Essentials

Video:

- 3Blue1Brown: Essence of Linear Algebra (full playlist, ~3 hours total) - This is non-negotiable, watch it all
- Khan Academy: Matrix multiplication (if needed for practice)

Reading:

- Start *ISLR* Chapter 3 (Linear Regression)

Hands-on:

- Implement matrix multiplication from scratch in numpy
- Understand broadcasting in numpy
- Solve a system of linear equations manually, then with numpy.linalg

Milestone: Can multiply matrices correctly and understand why shape $(m,n) \times (n,p) = (m,p)$.

Week 6: Linear Algebra for ML

Video:

- StatQuest: Linear Regression
- StatQuest: R-squared
- 3Blue1Brown: Eigenvectors and eigenvalues (conceptual understanding)

Reading:

- ISLR Chapter 3 continued

Hands-on:

- Implement simple linear regression from scratch using only numpy (normal equation)
- Compare results to sklearn LinearRegression
- Understand what's happening "under the hood"

Milestone: Can explain why linear regression is a linear algebra problem (solving for coefficients).

Week 7: sklearn Deep Dive - Preprocessing

Video:

- Any sklearn tutorial (DataCamp has good ones, or YouTube)
- Focus specifically on: Pipeline, StandardScaler, OneHotEncoder, train_test_split

Reading:

- sklearn documentation (yes, really read it)
- *Hands-On ML* Chapter 2

Hands-on:

- Master the fit/transform paradigm - create 10 examples
- Build preprocessing pipelines with ColumnTransformer
- Handle mixed data types (numerical + categorical)

Critical practice:

```python
# You should be able to write this pattern in your sleep:
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.model_selection import train_test_split

# Build a full preprocessing + model pipeline
# Understand WHY you only fit on training data
```

Milestone: Can build a complete sklearn pipeline from raw data to predictions without googling.

Week 8: sklearn Deep Dive - Model Selection

Video:

- StatQuest: Cross Validation
- StatQuest: ROC and AUC
- StatQuest: Precision, Recall, F1

Reading:

- ISLR Chapter 5 (Resampling Methods)
- *Hands-On ML* Chapter 3

Hands-on:

- Implement cross-validation manually, then use sklearn's cross_val_score
- Create confusion matrices, calculate precision/recall/F1 manually
- Build ROC curves for binary classification

Critical understanding:

- When accuracy is misleading (imbalanced classes!)
- Precision vs. recall trade-off
- When to use which metric

Milestone: Can explain why 99% accuracy on fraud detection is useless.

Phase 2: Core ML Algorithms (Weeks 9-16)

Now we go deep on the algorithms. The goal is not just to use them, but to understand *when*, *why*, and *how* they work.

Month 3: Supervised Learning Deep Dive

Week 9: Linear Models

Video:

- StatQuest: Logistic Regression (full series)

- StatQuest: Regularization (Ridge/Lasso/Elastic Net)

Reading:

- ISLR Chapter 4 (Classification)
- ISLR Chapter 6 (Regularization)

Hands-on:

- Implement logistic regression gradient descent from scratch
- Compare L1 vs L2 regularization effects on coefficients
- Use regularization to do feature selection

Milestone: Can explain why Lasso produces sparse coefficients but Ridge doesn't.

Week 10: Tree-Based Models

Video:

- StatQuest: Decision Trees
- StatQuest: Random Forests
- StatQuest: AdaBoost
- StatQuest: Gradient Boost

Reading:

- ISLR Chapter 8 (Tree-Based Methods)
- *Hands-On ML* Chapter 7

Hands-on:

- Train decision trees with varying depths, visualize overfitting
- Compare Random Forest vs. single tree
- Understand feature importances

Milestone: Can explain the difference between bagging (Random Forest) and boosting (XGBoost).

Week 11: Gradient Boosting Mastery (XGBoost, LightGBM, CatBoost)

Video:

- StatQuest: XGBoost series (all parts)
- YouTube tutorials on LightGBM, CatBoost

Reading:

- XGBoost documentation
- *Hands-On ML* gradient boosting section

Hands-on:

- Kaggle competition: Join a tabular data competition
- Experiment with hyperparameters: learning_rate, max_depth, n_estimators
- Learn to use early stopping

Milestone: Can train XGBoost with proper cross-validation and early stopping.

Week 12: Feature Engineering

Video:

- Kaggle's Feature Engineering course (free)
- Any advanced feature engineering tutorials

Reading:

- *Hands-On ML* relevant sections
- Kaggle competition winner write-ups (goldmine of techniques)

Hands-on:

- Create time-based features (relevant to your transaction work!)
- Aggregation features
- Encoding techniques for high-cardinality categoricals
- Target encoding (with proper cross-validation to avoid leakage!)

Milestone: Can take a raw dataset and engineer 10+ meaningful features.

Month 4: Evaluation, Tuning & Unsupervised Learning

Week 13: Hyperparameter Tuning & Model Selection

Video:

- GridSearchCV, RandomizedSearchCV tutorials
- Optuna tutorials (modern hyperparameter optimization)

Reading:

- sklearn documentation on model selection
- *Hands-On ML* Chapter 2-3 sections on tuning

Hands-on:

- Build a full ML pipeline with cross-validated hyperparameter search
- Compare GridSearch vs RandomSearch vs Optuna
- Learn to avoid overfitting during tuning (nested cross-validation)

Milestone: Can set up a proper, non-leaking hyperparameter search.

Week 14: Unsupervised Learning - Clustering

Video:

- StatQuest: K-means clustering
- StatQuest: Hierarchical clustering
- StatQuest: DBSCAN

Reading:

- ISLR Chapter 12
- *Hands-On ML* Chapter 9

Hands-on:

- Cluster customer transaction data (relevant to your work!)
- Evaluate clusters: silhouette score, elbow method
- Visualize with PCA/t-SNE

Milestone: Can explain when to use K-means vs. DBSCAN.

Week 15: Dimensionality Reduction

Video:

- StatQuest: PCA (main ideas, step-by-step)
- StatQuest: t-SNE
- 3Blue1Brown: Eigenvectors (review)

Reading:

- ISLR PCA section
- *Hands-On ML* Chapter 8

Hands-on:

- Implement PCA from scratch using eigendecomposition
- Use PCA for visualization and as preprocessing
- Understand explained variance ratio

Milestone: Can explain what principal components actually are.

Week 16: Model Interpretation & Explainability

Video:

- SHAP tutorials
- LIME tutorials

Reading:

- Christoph Molnar's "Interpretable Machine Learning" (free online)

Hands-on:

- Use SHAP to explain XGBoost predictions
- Create feature importance plots
- Understand partial dependence plots

Milestone: Can explain to a non-technical stakeholder why a model made a specific prediction.

Phase 3: Deep Learning & Advanced Topics (Weeks 17-20)

Week 17: Neural Network Fundamentals

Video:

- 3Blue1Brown: Neural Networks (full series)
- StatQuest: Neural Networks
- fast.ai Lesson 1-2

Reading:

- *Hands-On ML* Chapter 10

Hands-on:

- Implement a simple neural network from scratch (forward/backward pass)
- Build the same network in Keras/PyTorch
- Understand backpropagation conceptually

Milestone: Can explain backpropagation without reading notes.

Week 18: Practical Deep Learning

Video:

- fast.ai Lessons 3-5
- PyTorch tutorials

Reading:

- *Hands-On ML* Chapters 11-12

Hands-on:

- Build a classifier on a real dataset (MNIST, then something harder)
- Experiment with architectures, dropout, batch normalization

- Use GPU (Google Colab is free)

Milestone: Can train a neural network and diagnose basic issues (overfitting, vanishing gradients).

Week 19: NLP or Computer Vision Taster (Pick One)

Depending on interest, spend a week going deeper in one area:

NLP Path:

- Transformers basics, Hugging Face library
- Text classification or sentiment analysis project

Computer Vision Path:

- CNNs, transfer learning
- Image classification project

Milestone: Complete one end-to-end project in your chosen area.

Week 20: MLOps Fundamentals

Video:

- MLOps tutorials (Made With ML is good)
- Docker basics if needed

Reading:

- *Designing Machine Learning Systems* by Chip Huyen (excellent book)

Hands-on:

- Containerize a model with Docker
- Create a simple API endpoint (FastAPI)
- Basic model monitoring concepts

Milestone: Can take a model from notebook to deployable API.

Phase 4: Projects & Interview Prep (Weeks 21-26)

This is where everything comes together.

Weeks 21-23: Portfolio Projects

Build 2-3 solid projects that demonstrate different skills:

Project 1: End-to-end ML on tabular data

- Take a messy real-world dataset
- Full pipeline: EDA → cleaning → feature engineering → modeling → evaluation
- Use proper cross-validation, handle imbalanced classes
- Create a clear README and presentation
- *Suggested:* Fraud detection, churn prediction, or something finance-related (leverages your domain)

Project 2: Something with more ML depth

- Time series forecasting, recommendation system, or NLP
- Shows you can go beyond basic classification

Project 3: Deployed model

- Take one of your models and deploy it
- Simple web app or API
- Shows you understand the full lifecycle

Weeks 24-25: Interview Preparation

Practice:

- StrataScratch for DS questions
- Mock interviews (find a study partner or use Pramp)
- Review your projects, be ready to explain every decision

Study:

- Common DS interview questions

- ML system design basics
- Review fundamentals (they WILL ask about bias-variance, regularization, evaluation metrics)

Create:

- Clean GitHub profile with your projects
- LinkedIn update
- 1-page portfolio summary

Week 26: Job Applications

- Start applying to DS roles
- Customize applications
- Prepare for technical screens

Weekly Schedule Template

Here's how to structure a typical week:

Monday-Thursday (3-4 hours each)

Time	Activity
17:00-18:00	Video content (StatQuest, courses)
18:00-18:30	Break/dinner
18:30-20:30	Hands-on coding/practice
Before bed	30-45 min book reading

Friday (lighter day, 2-3 hours)

- Review week's concepts
- Light practice or reading
- Plan weekend project work

Saturday (4-5 hours)

- Project work
- Kaggle competitions
- Longer coding sessions

Sunday (3-4 hours or rest)

- Finish weekly milestone
- Preview next week's topics
- Rest if needed

Monthly Checkpoints

Use these to assess progress:

End of Month 1:

- [ ] Can explain p-values, confidence intervals, hypothesis testing
- [ ] Completed stats notebook with real examples
- [ ] Finished *Naked Statistics*

End of Month 2:

- [ ] Comfortable with matrix operations
- [ ] Can build sklearn pipelines without googling
- [ ] Understand evaluation metrics deeply

End of Month 3:

- [ ] Understand all major supervised algorithms
- [ ] Completed first Kaggle competition
- [ ] Can explain bias-variance tradeoff

End of Month 4:

- [ ] Comfortable with unsupervised learning
- [ ] Can use SHAP/model interpretation tools
- [ ] Started *Hands-On ML* book

End of Month 5:

- [ ] Basic neural network understanding
- [ ] Completed deep learning project

- [ ] Understand MLOps basics

End of Month 6:

- [ ] 2-3 portfolio projects complete
- [ ] GitHub profile polished
- [ ] Ready for interviews

Books Reading Order

Bedtime/conceptual (in order):

1. *Naked Statistics* - Month 1
2. *The Hundred-Page Machine Learning Book* - Month 2
3. *Approaching (Almost) Any Machine Learning Problem* - Month 3-4

Deep study (parallel with curriculum):

1. *ISLR* - Throughout, following the chapter guide above
2. *Hands-On Machine Learning* - Month 3 onwards
3. *Designing Machine Learning Systems* - Month 5

Final Notes

On consistency: 4 focused hours beats 6 distracted hours. If you're tired, do 2 good hours instead of 5 mediocre ones.

On frustration: You WILL feel stupid sometimes. Linear algebra will hurt. Backpropagation won't click immediately. This is normal. Push through.

On your advantage: You have something many DS candidates don't: 8 months of real data work. You understand messy data, stakeholders, and business context. Lean into this.

On the goal: By June, you won't know everything. But you'll have:

- Solid fundamentals you actually understand
- Practical experience building models
- Portfolio proof of your skills
- Confidence to tackle DS interviews

Start date: This week

First milestone: End of Week 1 - complete distributions content + start *Naked Statistics*

Good luck, Mads. You've got this.