Machine Learning

Complete Week-by-Week Roadmap

Guidexia Learning Platform

36	12	250+	9
Weeks	Phases	Skills	Months

www.guidexia.com

This is your complete, week-by-week Machine Learning roadmap — 12 phases across 36 weeks (9 months). Starting from Python and mathematics, you progress through data cleaning, feature engineering, supervised and unsupervised learning, deep learning with PyTorch, NLP with transformers, time-series forecasting, MLOps and production deployment, finishing with a full portfolio and career launch.

ROADMAP OVERVIEW — 12 PHASES | 36 WEEKS | 9 MONTHS

#	Phase	Timeline	Key Skills
01	Python & Mathematics for ML	Month 1	NumPy, Pandas, Linear Algebra, Statistics
02	Data Collection, Cleaning & EDA	Month 1-2	Data Cleaning, EDA, pandas-profiling
03	Feature Engineering & Preprocessing	Month 2	Scaling, Encoding, Feature Selection
04	Supervised Learning — Regression	Month 2-3	Linear Regression, Ridge, Lasso, XGBoost
05	Supervised Learning — Classification	Month 3	Logistic Regression, SVM, AUC-ROC, SMOTE
06	Unsupervised Learning	Month 4	K-Means, DBSCAN, PCA, t-SNE, UMAP
07	Model Evaluation, Pipelines & MLflow	Month 4-5	Scikit-learn Pipeline, MLflow, DVC, CV
08	Deep Learning with PyTorch	Month 5-6	PyTorch, CNN, Transfer Learning, W&B;
09	Natural Language Processing	Month 6-7	TF-IDF, BERT, Fine-tuning, Hugging Face
10	Time-Series & Advanced Topics	Month 7	ARIMA, Prophet, Anomaly, Recommenders
11	MLOps & Production Deployment	Month 8	FastAPI, Docker, SageMaker, Evidently AI

12	Specialisation, Portfolio & Career Launch	Month 8-9	Portfolio, Kaggle, ML Interviews, Career

CORE STACK	Python | NumPy | Pandas | Scikit-learn | PyTorch | Hugging Face | MLflow | FastAPI | Docker | AWS SageMaker

MONTH 01 Python + Math + Data Collection

1 Python & Mathematics for ML

Weeks 1-3 (Month 1)

PHASE 01

Machine learning is built on Python and mathematics. This phase gives you the exact Python skills and mathematical intuition — linear algebra, calculus, statistics — that ML engineers use every single day, taught through a lens of practical application.

WEEK Weeks 1-2 — Python for Machine Learning

Python ML Essentials	Python Patterns for ML
* NumPy: arrays, broadcasting, vectorisation	* List/dict comprehensions
* Pandas: DataFrames, groupby, merge	* Functional: map, filter, reduce
* Matplotlib & Seaborn visualisation	* File I/O: CSV, JSON, Parquet
* Scikit-learn first look	* Virtual environments & requirements.txt

WEEK Week 3 — Mathematics for ML

Linear Algebra	Calculus & Statistics
* Vectors, matrices, dot products	* Derivatives & gradient intuition
* Matrix multiplication & transpose	* Chain rule for backpropagation
* Eigenvalues & eigenvectors (intuition)	* Probability distributions
* PCA preview	* Mean, variance, covariance

SKILLS GAINED

Python	NumPy	Pandas	Matplotlib	Seaborn
Scikit-learn	Vectorisation	Broadcasting	Linear Algebra	Dot Products
Eigenvalues	Derivatives	Gradient	Probability	Statistics

Outcome:
You have the Python and mathematical foundations to read ML papers, understand algorithm derivations, and manipulate data at scale.

2 Data Collection, Cleaning & EDA

Weeks 4-5 (Month 1-2)

PHASE 02

Real-world ML starts with messy data. This phase teaches you to collect, clean, explore, and prepare datasets — the skill that occupies 80% of every ML engineer's time and determines whether models succeed or fail.

WEEK Week 4 — Data Collection & Cleaning

Data Sources	Data Cleaning
* Loading CSV, Excel, JSON, SQL databases	* Handling missing values: drop, impute, KNN
* Web scraping with BeautifulSoup	* Duplicate detection & removal
* API data collection	* Outlier detection: IQR, Z-score
* Handling large datasets (chunking)	* Data type conversion

WEEK Week 5 — Exploratory Data Analysis

EDA Techniques	Feature Understanding
* Univariate analysis: distributions	* Skewness & kurtosis
* Bivariate: correlation matrix, scatter plots	* Class imbalance detection
* Multivariate: heatmaps, pair plots	* Feature relationships
* Summary statistics	* EDA report with pandas-profiling

SKILLS GAINED

Data Collection	Web Scraping	SQL Data	Missing Values	Imputation
KNN Imputation	Outlier Detection	IQR	Z-score	EDA
Correlation Matrix	Pair Plots	Heatmaps	Class Imbalance	pandas-profiling

Outcome:
You can take any raw dataset, clean it systematically, detect quality issues, and produce a comprehensive EDA report with actionable insights.

MONTH 02 Feature Engineering + Regression

3 Feature Engineering & Preprocessing

Weeks 6-7 (Month 2)

PHASE 03

Feature engineering is where domain knowledge meets machine learning. The features you create often matter more than the algorithm you choose. This phase teaches you to transform raw data into powerful model inputs.

WEEK Week 6 — Feature Engineering

Numerical Features	Categorical Features
* Scaling: StandardScaler, MinMaxScaler, RobustScaler	* One-hot encoding & ordinal encoding
* Log & Box-Cox transformation	* Target encoding
* Polynomial features	* Frequency encoding
* Binning & discretisation	* Handling high-cardinality

WEEK Week 7 — Advanced Feature Engineering

Time-Series Features	Feature Selection
* Date/time decomposition	* Filter methods: correlation, chi-squared
* Lag features & rolling windows	* Wrapper: RFE
* Cyclic encoding (sine/cosine)	* Embedded: Lasso
* Trend & seasonality	* Variance threshold

SKILLS GAINED

StandardScaler	MinMaxScaler	RobustScaler	Log Transform	Box-Cox
Polynomial Features	Binning	One-Hot Encoding	Target Encoding	Frequency Encoding
Lag Features	Rolling Windows	Cyclic Encoding	RFE	Lasso
Feature Selection				

Outcome:

You can engineer a complete feature set from raw data, apply proper scaling, encode categoricals, and select the most predictive features.

4 Supervised Learning — Regression

Weeks 8-10 (Month 2-3)

PHASE 04

Regression is the foundation of predictive modelling. This phase covers every regression technique — from simple linear models to gradient boosting — with deep focus on when to use each and how to evaluate them properly.

WEEK Week 8 — Linear Models

Linear Regression	Regularised Regression
* Ordinary Least Squares (OLS)	* Ridge (L2) regression
* Assumptions & diagnostics	* Lasso (L1) regression
* Residual analysis	* ElasticNet
* Interpretation of coefficients	* Choosing regularisation strength

WEEK Week 9 — Tree-Based Regression

Decision Trees	Ensemble Methods
* CART algorithm	* Random Forest Regressor
* Splitting criteria: MSE, MAE	* Gradient Boosting (GBM)
* Pruning & max_depth	* XGBoost & LightGBM
* Overfitting in trees	* Feature importance

WEEK Week 10 — Regression Evaluation & Tuning

Regression Metrics	Hyperparameter Tuning
* MAE, MSE, RMSE, R-squared	* GridSearchCV
* MAPE & SMAPE	* RandomizedSearchCV
* Adjusted R-squared	* Optuna framework
* Residual plots	* Cross-validation strategies

SKILLS GAINED

Linear Regression	OLS	Ridge	Lasso	ElasticNet
Decision Trees	Random Forest	GBM	XGBoost	LightGBM
Feature Importance	MAE	RMSE	R-squared	GridSearchCV
Optuna	Cross-Validation			

Outcome:
You can build, evaluate, tune, and compare any regression model and select the best one for a given business problem with proper validation.

MONTH 03 Classification

5 Supervised Learning — Classification

Weeks 11-13 (Month 3)

PHASE 05

Classification is the most common ML task in industry — fraud detection, churn prediction, spam filtering, medical diagnosis. This phase makes you expert in all major classification algorithms and their evaluation.

WEEK Week 11 — Core Classifiers

Logistic Regression	K-Nearest Neighbours & Naive Bayes
* Sigmoid function & probability	* KNN: distance metrics
* Binary vs multi-class	* Choosing K
* Threshold tuning	* Gaussian Naive Bayes
* Log-loss interpretation	* Bernoulli & Multinomial NB

WEEK Week 12 — Advanced Classifiers

Support Vector Machines	Tree Ensembles
* Maximum margin classifier	* Random Forest Classifier
* Kernel trick (RBF, polynomial)	* Gradient Boosting Classifier
* C & gamma hyperparameters	* XGBoost & CatBoost
* SVM for multi-class	* Stacking & voting classifiers

WEEK Week 13 — Classification Evaluation

Metrics	Imbalanced Classes
* Confusion matrix	* SMOTE & ADASYN oversampling
* Precision, Recall, F1-score	* Undersampling strategies
* AUC-ROC & AUC-PR	* Class weight adjustment
* Matthews Correlation Coefficient	* Threshold optimisation

SKILLS GAINED

Logistic Regression	KNN	Naive Bayes	SVM	Kernel Trick
Random Forest	XGBoost	CatBoost	Stacking	Confusion Matrix
Precision	Recall	F1	AUC-ROC	AUC-PR
SMOTE	ADASYN	Class Weights		

Outcome:
You can train and evaluate any classifier, handle class imbalance, optimise decision thresholds, and report results correctly to business stakeholders.

MONTH 04 Unsupervised + Pipelines + MLflow

6 Unsupervised Learning

Weeks 14-15 (Month 4)

PHASE 06

Not all data has labels. Unsupervised learning finds hidden structure — customer segments, anomalies, compressed representations — without supervision. These techniques are increasingly important in modern AI systems.

WEEK Week 14 — Clustering

K-Means & Variants	Advanced Clustering
* K-Means algorithm deep dive	* DBSCAN: density-based clustering
* Elbow method & silhouette score	* Hierarchical clustering & dendrograms
* K-Means++ initialisation	* Gaussian Mixture Models (GMM)
* Mini-batch K-Means	* Cluster evaluation metrics

WEEK Week 15 — Dimensionality Reduction

Linear Methods	Non-Linear Methods
* PCA: explained variance ratio	* t-SNE: perplexity & learning rate
* Choosing n_components	* UMAP for visualisation
* Kernel PCA	* Autoencoders preview
* Factor Analysis	* When to use each method

SKILLS GAINED

K-Means	Elbow Method	Silhouette Score	DBSCAN	Hierarchical Clustering
Dendrograms	GMM	PCA	Explained Variance	Kernel PCA
t-SNE	UMAP	Dimensionality Reduction	Cluster Evaluation	

Outcome:
You can segment data with clustering, visualise high-dimensional datasets with t-SNE/UMAP, and reduce dimensions for downstream ML tasks.

7 Model Evaluation, Pipelines & MLflow

Weeks 16-18 (Month 4-5)

PHASE 07

A model in a notebook is not production ML. This phase teaches you to build reproducible pipelines, track experiments professionally, and evaluate models rigorously — the skills that distinguish ML engineers from notebook explorers.

WEEK Week 16 — Scikit-learn Pipelines

Pipeline Architecture

* Pipeline object: steps & named_steps
* ColumnTransformer for mixed types
* FunctionTransformer
* Custom transformer classes

Pipeline Best Practices

* Preventing data leakage
* Persisting pipelines (joblib, pickle)
* Pipeline with cross-validation
* Production-ready pipeline design

WEEK Week 17 — Experiment Tracking with MLflow

MLflow Core

* mlflow.log_param & log_metric
* Logging artefacts (models, plots)
* MLflow runs & experiments
* Model registry

Reproducibility

* Setting random seeds
* Versioning datasets with DVC
* Environment management
* Reproducible training scripts

WEEK Week 18 — Advanced Model Evaluation

Evaluation Frameworks

* Nested cross-validation
* Time-series cross-validation
* Stratified k-fold
* Bootstrap confidence intervals

Model Comparison

* Statistical tests: Wilcoxon, McNemar
* Learning curves
* Bias-variance tradeoff analysis
* Model selection criteria (AIC, BIC)

SKILLS GAINED

Scikit-learn Pipeline	ColumnTransformer	Custom Transformer	Data Leakage Prevention	MLflow
Log Params	Log Metrics	Model Registry	DVC	Nested CV
Time-Series CV	Stratified K-Fold	Bootstrap CI	Learning Curves	Bias-Variance

Outcome:

You can build production pipelines, track experiments with MLflow, version data with DVC, and evaluate models with statistically rigorous methods.

MONTH 05 Deep Learning with PyTorch

8 Deep Learning with PyTorch

Weeks 19-22 (Month 5-6)

PHASE 08

Deep learning powers image recognition, NLP, and generative AI. This phase builds genuine PyTorch expertise — from tensors to training loops to CNNs — understanding what happens inside the black box.

WEEK Week 19 — PyTorch Fundamentals

Tensors & Autograd	Neural Network Basics
* Tensor creation & operations	* nn.Module architecture
* Broadcasting in PyTorch	* Linear layers & activations
* autograd & computational graphs	* Forward pass implementation
* requires_grad & backward()	* Loss functions: CE, MSE, BCE

WEEK Week 20 — Training Deep Networks

Training Loop	Regularisation
* DataLoader & Dataset classes	* Dropout layers
* Training & validation loop	* Batch Normalisation
* Optimisers: SGD, Adam, AdamW	* Weight decay (L2)
* Learning rate schedulers	* Early stopping

WEEK Week 21 — CNNs — Computer Vision

CNN Architecture	Transfer Learning
* Conv2d: kernels, stride, padding	* Pretrained models (ResNet, VGG, EfficientNet)
* Pooling layers	* Feature extraction vs fine-tuning
* Receptive field concept	* torchvision transforms
* BatchNorm & activation order	* Custom dataset for images

WEEK Week 22 — Training at Scale

GPU Training	Experiment Tracking
* Moving tensors to CUDA	* Weights & Biases (W&B;) integration
* Mixed precision (torch.amp)	* TensorBoard logging
* DataParallel & DistributedDataParallel	* Model checkpointing
* Gradient accumulation	* Early stopping patterns

SKILLS GAINED

PyTorch	Tensors	Autograd	nn.Module	DataLoader
Adam	AdamW	LR Scheduler	Dropout	BatchNorm
CNN	Conv2d	ResNet	EfficientNet	Transfer Learning
Fine-Tuning	CUDA	Mixed Precision	W&B;	TensorBoard

Outcome:
You can build, train and debug neural networks in PyTorch, implement CNNs with transfer learning, and train efficiently on GPU with experiment tracking.

MONTH 06 NLP + Transformers

9 Natural Language Processing

Weeks 23-25 (Month 6-7)

PHASE 09

Language is the primary interface of AI in 2024-2025. This phase takes you from classical NLP through transformers and Hugging Face — the skills that power chatbots, summarisers, classifiers, and LLM applications.

WEEK Week 23 — Classical NLP

Text Preprocessing	Feature Extraction
* Tokenisation: word, subword, BPE	* Bag-of-Words & TF-IDF
* Stemming, lemmatisation	* Word2Vec & GloVe embeddings
* Stop word removal	* FastText
* Regular expressions for text	* N-grams

WEEK Week 24 — Transformers & Hugging Face

Transformer Architecture	Hugging Face
* Self-attention mechanism	* from_pretrained API
* Multi-head attention	* Tokenizer & model pipeline
* Positional encodings	* AutoModel & AutoTokenizer
* Encoder vs decoder architectures	* Inference with transformers

WEEK Week 25 — Fine-tuning & NLP Applications

Fine-tuning BERT	NLP Tasks
* Text classification task	* Named Entity Recognition (NER)
* Trainer API	* Question Answering
* Custom training loop	* Text Summarisation (T5/BART)
* Evaluation: accuracy, F1	* Sentiment analysis

SKILLS GAINED

Tokenisation	BPE	TF-IDF	Word2Vec	GloVe
Transformers	Self-Attention	Multi-Head Attention	Positional Encoding	Hugging Face
BERT	Fine-tuning	Trainer API	NER	Question Answering
Summarisation	Sentiment Analysis			

Outcome:
You can fine-tune transformer models for any NLP task, work with the Hugging Face ecosystem, and build text classification and extraction systems.

MONTH 07 Time-Series + Advanced Topics

10 Time-Series & Advanced Topics

Weeks 26-28 (Month 7)

PHASE 10

Time-series forecasting drives billions in decisions across finance, retail, and operations. This phase also covers reinforcement learning, anomaly detection, and recommendations — rounding out your ML expertise.

WEEK Week 26 — Time-Series Forecasting

Classical Methods	ML for Forecasting
* ARIMA: AR, I, MA components	* Prophet by Meta
* Seasonal ARIMA (SARIMA)	* LightGBM for time-series
* Exponential smoothing (Holt-Winters)	* Feature engineering for TS
* Stationarity & ADF test	* Evaluation: MAPE, RMSE, MAE

WEEK Week 27 — Anomaly Detection & Recommender Systems

Anomaly Detection	Recommender Systems
* Isolation Forest	* Collaborative filtering
* One-Class SVM	* Content-based filtering
* Autoencoder for anomaly detection	* Matrix factorisation (SVD)
* Statistical process control	* Hybrid recommender systems

WEEK Week 28 — Reinforcement Learning Intro

RL Fundamentals	Deep RL
* Agent, environment, reward, policy	* Deep Q-Network (DQN)
* Markov Decision Process (MDP)	* Policy gradient methods
* Q-learning algorithm	* Gymnasium (OpenAI Gym)
* Epsilon-greedy exploration	* Stable-Baselines3

SKILLS GAINED

ARIMA	SARIMA	Holt-Winters	Stationarity	Prophet
LightGBM TS	MAPE	Isolation Forest	One-Class SVM	Autoencoder Anomaly
Collaborative Filtering	Matrix Factorisation	SVD	MDP	Q-Learning
DQN	Policy Gradient	Gymnasium		

Outcome:

You can build time-series forecasting systems, detect anomalies, create recommender systems, and understand reinforcement learning fundamentals.

MONTH 08 MLOps + Deployment + Portfolio

11 MLOps & Production Deployment

Weeks 29-32 (Month 8)

PHASE 11

A model that nobody uses creates zero value. MLOps is the discipline that bridges research and production — the skill that makes you irreplaceable in any ML team and earns the highest salaries.

WEEK Week 29 — Model Serving & APIs

FastAPI for ML	Serving Frameworks
* Creating prediction endpoints	* BentoML for model packaging
* Request/response schemas (Pydantic)	* TorchServe for PyTorch
* Async inference	* Triton Inference Server
* Health checks	* Batch vs real-time inference

WEEK Week 30 — Containerisation & Cloud

Docker for ML	Cloud Deployment
* Dockerfile for ML services	* AWS SageMaker endpoint
* Docker Compose for multi-service	* GCP Vertex AI
* Base images: python-slim, CUDA	* Azure ML
* Multi-stage builds	* Modal for serverless GPU

WEEK Week 31 — CI/CD & Model Monitoring

CI/CD Pipelines	Model Monitoring
* GitHub Actions for ML	* Evidently AI for data drift
* Automated testing of ML code	* Model performance degradation
* Model validation in CI	* Prediction monitoring
* Automated retraining triggers	* Alerting pipelines

WEEK Week 32 — Feature Stores & Orchestration

Feature Stores	Pipeline Orchestration
* Feast feature store	* Airflow for ML pipelines
* Online vs offline stores	* Prefect as alternative
* Feature versioning	* Kubeflow Pipelines
* Feature serving latency	* End-to-end pipeline design

SKILLS GAINED

FastAPI	Pydantic	Async Inference	BentoML	TorchServe
Triton	Docker	Dockerfile	SageMaker	Vertex AI
Azure ML	Modal	GitHub Actions CI	Evidently AI	Data Drift
Model Monitoring	Feast	Airflow	Prefect	Kubeflow

Outcome:
You can deploy ML models as production APIs, containerise them, monitor for drift, and build end-to-end orchestrated ML pipelines in the cloud.

12 Specialisation, Portfolio & Career Launch

Weeks 33-36 (Month 8-9)

PHASE 12

This final phase builds your ML portfolio, prepares you for interviews, and establishes your technical reputation — turning 8 months of learning into a career in machine learning.

WEEK Weeks 33-34 — Capstone Projects & Portfolio

Capstone ML Project	Portfolio Assets
* End-to-end: EDA → features → model → API	* 3 ML case studies on GitHub
* Deployed on cloud with monitoring	* Kaggle profile with public notebooks
* MLflow experiment tracking	* Technical blog posts (Medium/Dev.to)
* GitHub with full documentation	* Hugging Face model card published

WEEK Weeks 35-36 — Interviews & Career Strategy

ML Interview Preparation	Career Launch
* ML fundamentals questions	* ML engineer vs data scientist vs researcher
* Statistics & probability	* Resume: quantify model impact
* Coding: arrays, strings, recursion	* LinkedIn: ML positioning
* System design for ML at scale	* Target companies & networking

SKILLS GAINED

Capstone Project	End-to-End ML	MLflow Portfolio	API Deployment	GitHub Portfolio
Kaggle	Hugging Face	Technical Writing	ML Interviews	Statistics Questions
ML System Design	Resume	LinkedIn	Networking	Career Strategy

Outcome:
You have a deployed ML project, published case studies, Kaggle & Hugging Face presence, and complete preparation to land ML engineer and data scientist roles.

SUCCESS TIPS & FREE RESOURCE DIRECTORY

Implement Everything From Scratch Once	Build linear regression, backpropagation, and k-means from scratch before using libraries. Understanding the math makes you 10x better with Scikit-learn.
Kaggle From Month 2 — Not Month 9	Enter competitions with the skills you have now. The discussions, kernels, and leaderboard teach more than any tutorial at the same stage.
Read One ML Paper Per Week From Month 4	Start with Attention Is All You Need, then XGBoost, then BERT. Papers give you the depth that separates engineers from practitioners.
MLflow Every Experiment — Always	Track every run from your first scikit-learn model. The habit of reproducibility is what makes you valuable in production ML teams.
Deploy Before Month 9	Get something live — even a simple FastAPI model on a free server. Deployed projects in your portfolio are worth more than 10 certificates.
Specialise in One Domain by Month 7	NLP, computer vision, time-series, or recommender systems. Deep domain knowledge combined with ML skills is what gets you the best roles.
Write About What You Learn	A Medium post explaining gradient descent in plain English, a Hugging Face model card, a GitHub README — these build your reputation as an expert.

FREE RESOURCES

Resource	URL	Best For
fast.ai	fast.ai	Best practical deep learning course
Hugging Face Learn	huggingface.co/learn	Transformers, NLP, diffusion — all free
StatQuest	youtube: StatQuest	Best statistics & ML explanations
Kaggle Learn	kaggle.com/learn	Free micro-courses + competitions
PyTorch Docs	pytorch.org/docs	Official PyTorch tutorials
Scikit-learn Docs	scikit-learn.org/stable	Best ML library documentation
CS229 Stanford	cs229.stanford.edu	Andrew Ng's ML course notes (free)
MLflow Docs	mlflow.org/docs	Experiment tracking & model registry
Andrej Karpathy	youtube: Andrej Karpathy	Build GPT/NN from scratch videos
Papers With Code	paperswithcode.com	ML papers + code implementations
Evidently AI Docs	docs.evidentlyai.com	Model & data monitoring
Guidexia	www.guidexia.com	Structured roadmaps, mentorship & community

In 9 months, you will build systems that learn. Start with curiosity, finish with models that ship.

www.guidexia.com