SEED

1-Month Machine Learning Internship

4 Weeks | 7+ Hours Daily | Foundational → Intermediate

Program Overview

Focus: "Algorithm Intuition" — Building logic from scratch before using high-level libraries. This program transitions students from basic Python coding to building full-stack ML components.

WEEK 1 Python & NLP Architecture

Objective: Master the data structures and logic behind how machines 'read' human language

Day 1	**Topics:** Advanced Lists/Dicts, List Comprehensions, Lambda Functions, Math module
	Activity: *Handle edge cases in text (line breaks, encoding errors)* ★ **Task: Advanced Word Frequency Engine – Count occurrences across 10+ files and export results to JSON**
Day 2	**Topics:** Regex for text cleaning, Stopword logic, N-grams
	Activity: *Build a class-based tokenizer* ★ **Task: SimpleTokenizer Class – Implement .clean(), .tokenize(), and .remove_stopwords() methods**
Day 3	**Topics:** os and glob modules, Dynamic file indexing
	Activity: *Build a global vocabulary from a directory of raw text* ★ **Task: Document-to-ID Mapper – Create a bi-directional mapping system (Word ↔ Index)**
Day 4	**Topics:** One-Hot vectors, Co-occurrence logic, Dot Product math
	Activity: *Manual implementation of Cosine Similarity*

★ **Task: Mini Embedding Engine – Input a word and find the 'top 3 most similar words' in the local corpus**

Day 5

Topics: Modular programming, Git basics

Activity: *Integrating all Week 1 components into a single workflow*

★ **Task: The Tiny NLP Engine – A CLI script that takes a folder of text and allows similarity queries**

WEEK 2 Data Intensive & Search Architecture

Objective: Master the 'Garbage In, Garbage Out' rule and build an Information Retrieval (IR) system

Day 1

Topics: Data acquisition from CSV/JSON, Basic Web Scraping (BeautifulSoup)

Activity: *Fetching and exploring real-world datasets*

★ **Task: The 'Dirty Data' Challenge – Clean a dataset with missing values, duplicates, and inconsistent date formats**

Day 2

Topics: Label Encoding vs. One-Hot Encoding, Feature Scaling (StandardScaler vs. MinMax)

Activity: *Transform raw data into ML-ready format*

★ **Task: Feature Engineering Lab – Transform a raw dataset into a clean numerical matrix ready for Scikit-Learn**

Day 3

Topics: Matplotlib & Seaborn basics, Identifying correlations and outliers

Activity: *Exploratory Data Analysis (EDA) on real datasets*

★ **Task: Insights Report – Generate 5 visual charts that explain 'Why' certain data points are related**

Day 4

Topics: Term-Incidence Matrices vs. Inverted Indexes, Boolean Retrieval

Activity: *Building a lookup table for instant search*

★ **Task: Inverted Index Builder – System that identifies which documents contain a keyword in O(1) time**

Day 5

Topics: Streamlit layout, State management, File uploaders

Activity: *Building a web-based search interface*

★ **Task: Final Micro Search Engine – A web app where users upload files and search with ranked results**

WEEK 3 Machine Learning Algorithms

Objective: Deep dive into Supervised, Unsupervised, and Reinforcement Learning

Day 1	**Topics:** Gradient Descent (Intuition), Mean Squared Error (MSE), Sigmoid function
	Activity: *Implement regression from scratch, then with Scikit-Learn* ★ **Task: GPA Predictor – Predict final student grades based on study hours and attendance**
Day 2	**Topics:** Entropy, Gini Impurity, Intro to Random Forests
	Activity: *Visualize decision tree splits on real data* ★ **Task: Loan Approval System – Classify applications into 'Approve' or 'Reject'**
Day 3	**Topics:** K-Means (Centroid movement), DBSCAN (Density-based clustering)
	Activity: *Compare clustering outputs on synthetic data* ★ **Task: Market Basket Analysis – Cluster local shop sales data to identify buying patterns**
Day 4	**Topics:** Principal Component Analysis (PCA), Variance explained
	Activity: *Apply PCA to a high-dimensional dataset and visualize results* ★ **Task: Data Compressor – Reduce a 20-feature dataset to 2D for visualization without losing core patterns**
Day 5	**Topics:** Agents, Environments, Rewards, Q-Learning
	Activity: *Build a simple grid-based RL environment* ★ **Task: Campus Navigation Simulation – Train an agent to find the shortest path from gate to classroom**

WEEK 4 Capstone Projects & Career Prep

Objective: Build, Document, and Present a complete ML project

Mon–Tue: Capstone Development

Students choose one of three specialization tracks:

Track 1	Track 2	Track 3
Predictive Analytics	**Customer Segmentation**	**Information Retrieval**
GPA/Success prediction using Supervised Learning models	Sales behavior clustering using Unsupervised Learning	Advanced Search system with Semantic Ranking

- Wednesday: Model Tuning & Evaluation — Focus on Precision/Recall, F1-Score, and Cross-Validation
- Thursday: Documentation & Portfolio — Write a professional README.md and host the project on GitHub
- Friday: Presentation & Demo — Deliver a 10-minute demo of the working system to stakeholders

Assessment Structure

Component	Weight	Criteria
Weekly Mini-Projects	**30%**	Code cleanliness & Logic accuracy
Micro Search Engine	**20%**	UI Usability & Retrieval Speed
ML Labs	**20%**	Proper use of Evaluation Metrics
Final Capstone	**30%**	Problem solving & Presentation

Program Philosophy: Every concept is first built from scratch to develop true understanding. High-level libraries are tools — not crutches. Students leave this program knowing the math, the code, and the craft.