ML Engineer

12 Week Roadmap

Step by step with Resources

By- @helloworld_avani

ML Engineer 12 Week Roadmap

✅ Week 1: Python Programming for ML

🔹 **Topics:**

- Python Installation (Anaconda or native)
- Variables, Data Types (int, float, str, bool)
- Operators (Arithmetic, Assignment, Logical, Comparison)
- Conditional Statements (`if`, `elif`, `else`)
- Loops (`for`, `while`, `break`, `continue`)
- Functions (parameters, return, lambda functions)
- Data Structures: List, Tuple, Set, Dictionary
- List Comprehensions
- File I/O and Basic Exception Handling

🔗 **Resources:**

- https://www.w3schools.com/python/
- https://www.coursera.org/specializations/python
- https://www.learnpython.org/

✅ Week 2: Python for Data Science

🔹 **Topics:**

- NumPy: Arrays, Array Indexing, Broadcasting, Math operations
- Pandas:
    - Series vs DataFrame
    - Reading/Writing CSV & Excel
    - DataFrame slicing, filtering, sorting
    - Missing values (`isnull()`, `dropna()`, `fillna()`)
    - GroupBy, Aggregation, Pivot Table
- Basic Data Cleaning
- Working with DateTime

🔗 **Resources:**

- https://www.kaggle.com/learn/python
- https://www.kaggle.com/learn/pandas
- https://numpy.org/doc/stable/user/quickstart.html

✅ Week 3: Statistics and Probability for ML

🔹 Topics:

- Descriptive Statistics:
    - Mean, Median, Mode, Range
    - Variance & Standard Deviation
- Probability Basics:
    - Independent vs Dependent Events
    - Bayes' Theorem, Conditional Probability
- Distributions:
    - Normal, Binomial, Poisson
- Central Limit Theorem
- Hypothesis Testing, p-values, t-test, z-test
- Confidence Intervals

🔗 Resources:

- https://www.khanacademy.org/math/statistics-probability
- https://www.youtube.com/user/joshstarmer (StatQuest)

✅ Week 4: Data Visualization

🔹 Topics:

- Matplotlib:
    - Line, Bar, Scatter, Histogram
    - Customization: labels, titles, colors
- Seaborn:
    - Distribution plots, Box plots, Violin plots
    - Heatmaps, Pairplots, Countplots
- Plotly for Interactive Plots (optional)

🔗 Resources:

- https://www.kaggle.com/learn/data-visualization
- https://seaborn.pydata.org/tutorial.html
- https://matplotlib.org/stable/gallery/index.html

✅ Week 5: Exploratory Data Analysis (EDA)

🔹 **Topics:**

- Understanding dataset structure
- Handling Missing Values, Duplicates
- Detecting and treating Outliers (Z-score, IQR)
- Feature Engineering:
    - Creating new features
    - Encoding categorical variables (One-hot, Label)
    - Scaling (MinMax, StandardScaler)
- Correlation Matrix
- Creating an EDA Report using `pandas-profiling` or `sweetviz`

🔗 **Resources:**

- https://www.kaggle.com/code/pmarcelino/comprehensive-data-exploration-with-python
- https://github.com/pandas-profiling/pandas-profiling
- https://pypi.org/project/sweetviz/

✅ Week 6: Machine Learning Basics

🔹 **Topics:**

- What is ML? Types of ML (Supervised, Unsupervised, Reinforcement)
- Workflow: Data ➜ Train-Test Split ➜ Model ➜ Evaluate
- Underfitting vs Overfitting
- Bias-Variance Tradeoff
- Evaluation Metrics:
    - Regression: MSE, MAE, RMSE, R^2
    - Classification: Accuracy, Precision, Recall, F1-Score

🔗 **Resources:**

- https://developers.google.com/machine-learning/crash-course
- https://www.kaggle.com/learn/intro-to-machine-learning
- https://scikit-learn.org/stable/

✅ Week 7: Regression Algorithms

🔹 **Topics:**

- Linear Regression (Single & Multi-variable)
- Polynomial Regression
- Ridge & Lasso Regression
- Model Evaluation: R^2, Adjusted R^2, RMSE
- Assumptions of Linear Regression
- Feature Scaling & Normalization

🔗 **Resources:**

- https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/
- https://www.youtube.com/watch?v=nk2CQITm_eo (StatQuest)
- https://scikit-learn.org/stable/modules/linear_model.html

✅ Week 8: Classification Algorithms

🔹 Topics:

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Decision Tree Classifier
- Random Forest
- Naive Bayes
- Support Vector Machine (SVM)
- Confusion Matrix, ROC-AUC, Precision-Recall Curve

🔗 Resources:

- https://www.kaggle.com/learn/intermediate-machine-learning
- https://www.youtube.com/@codebasics
- https://scikit-learn.org/stable/user_guide.html

✅ Week 9: Unsupervised Learning

🔹 **Topics:**

- Clustering:
  - K-Means: Elbow method, Silhouette Score
  - Hierarchical Clustering: Dendrogram
- Dimensionality Reduction:
  - Principal Component Analysis (PCA)
  - t-SNE (visualization only)

🔗 **Resources:**

- https://www.youtube.com/user/Simplilearn
- https://scikit-learn.org/stable/modules/clustering.html
- https://towardsdatascience.com/k-means-clustering-explained-4528df86a120

✅ Week 10: Feature Engineering & Hyperparameter Tuning

🔹 **Topics:**

- Feature Engineering:
  - Binning, Interaction Terms, Transformations (log, square root)
- Feature Selection:
  - Correlation, Chi-Square, Recursive Feature Elimination
- Model Tuning:
  - Grid Search CV, Randomized Search CV
  - Cross Validation (K-Fold)

🔗 **Resources:**

- https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/
- https://scikit-learn.org/stable/modules/grid_search.html
- https://www.kaggle.com/learn/feature-engineering

✅ Week 11: Capstone Projects

🔹 Project Ideas:

- House Price Prediction (Regression)
- Credit Card Fraud Detection (Classification)
- Customer Segmentation (Clustering)
- Email/SMS Spam Classifier (NLP)
- Breast Cancer Detection (Classification)
- Wine Quality Prediction

🔹 Practice Platforms:

- Kaggle Competitions
- GitHub Repositories
- Personal Portfolio Website

🔗 Resources:

- https://www.kaggle.com/datasets
- https://github.com/topics/machine-learning-projects
- https://ml-showcase.streamlit.app/

✅ Week 12: Job Preparation & Deployment

🔹 **Topics:**

- Git & GitHub Basics (Version Control)
- Create a Portfolio with ReadMe Files
- Resume Building with Projects & Skills
- LinkedIn Optimization & Networking
- Learn Model Deployment:
    - Flask / Streamlit
    - Deploy on Render / HuggingFace / Heroku

🔹 **Optional Tools:**

- Docker basics
- Streamlit UI customization

🔗 **Resources:**

- https://www.overleaf.com/latex/templates/data-science-resume/zcdmpfxvpdps
- https://streamlit.io/
- https://github.com/heroku/python-getting-started
- https://www.linkedin.com/jobs/

Comment 'ML'
for Pdf version of this post