iSCALE

AI Engineer Roadmap

Here's a curated roadmap tailored for absolute beginners aiming to venture into AI Engineer irrespective of their prior coding or computer science experience. This roadmap not only provides free resources for acquiring technical skills but also emphasizes the development of essential soft skills crucial for success in the field.

Total Duration: **6-8 Months**

3 hours in Tool Skills + **1 hour** in Core Skills = **4 hours** study Every Day

<mark>Module 0– Do your Research</mark>

<mark>Module 1 - Python</mark>

1. **The Basics of Python Programming**
2. **Variables and Operators in Python**
3. **Built-in Functions of Python**
4. **Control Flow Statements**
5. **Concepts of Loops**
6. **User-Defined Functions in Python**
7. **Data Structure - Strings in Python**
8. **Data Structure – List in Python**
9. **Data Structure – Tuples in Python**
10. **Data Structure – Dictionary in Python**
11. **Data Structure – Set in Python**
12. **Object Oriented Programming**
13. **Files Handling**
14. **Exception Handling**

Free Resources - http://tinyurl.com/38bx6fkk

<mark>Module 2 - Python Libraries</mark>

- **Numpy**

- **Pandas**

- **Matplotlib**

- **Seaborn**

- **Plotly**

LinkedIn - Core Skill
Create a professional-looking LinkedIn profile.
- Have a clear profile picture and banner image.
- Add tags such as: Open to work etc.

Business Fundamentals - Soft Skill
- Learn business concepts from YT Case Studies
- Example: Sales Case study asked in shopify : http://tinyurl.com/4a5f58yd

Module 4 - Math and Statistics for Data Science
- Basics: Descriptive vs inferential statistics, continuous vs discrete data, nominal vs ordinal data
- Basic plots: Histograms, pie charts, bar charts, scatter plot etc.
- Measures of central tendency: mean, median, mode
- Measures of dispersion: variance, standard deviation
- Probability basics
- Distributions: Normal distribution
- Correlation and covariance
- Central limit theorem
- Hypothesis testing: p value, confidence interval, type 1 vs type 2 error, Z test, t test, ANOVA

Module 5 Feature Engineering
Unit 1
- AI vs ML vs DL vs DS
- Supervised, Unsupervised Reinforcement Learning
- Train, Test Validation
- Variance, Bias, Overfitting Underfitting

Unit 2
- Handling Missing Values
- Handling Imbalanced Dataset
- SMOTE
- Data Interpolation
- Handling Outliers

Unit 3

- Feature Extraction
- Feature Scaling Normalization
- Normalization Min Max Scaling
- Unit Vectors Feature Scaling
- PCA

Unit 4
- Data Encoding
- Nominal or One Hot Encoding

Unit 5
- Label Ordinal Encoder
- Target Guided Ordinal Encoding

Unit 6
- Covariance And Correlation
- Covariance And Correlation with Python

Free Resource - http://tinyurl.com/4erv983d

Module 6 Exploratory Data Analysis

Unit 1
- EDA With Red Wine Data
- EDA Student Performance Indicator
- EDA Forest Fires

Unit 2
- EDA Flight Price
- EDA Google Playstore

Module 7 Machine Learning – I Supervised

Part 1 – Regression

Unit 1
- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Linear Regression

Unit 2
- R squared Adjusted R squared
- MSE,MAE RMSE
- Simple Linear Regression With Python
- Multiple Linear Regression

Unit 3
- Ridge Regression
- Lasso Regression

Unit 4
- Elastic Net Regression
- EDA with Algerian Forest Fire

- Model Training for Ridge, Lasso Elastic Net

Part 2 – Logistics Regression
- Logistic Regression Indepth Intuition
- Logistic Regression with Regularization
- Performance Metrics Confusion Matrix, Accuracy, Precision Recall
- Cross Validation Types
- Hyperparameter Tuning
- Logistic Regression Implementation
- Logistic Regression Multiclass Classification

Part 3 – Decision Tree
- Decision Tree Classifier Intuition
- Decision Tree For Numerical Split
- Post Pruning And Prepruning Decision Trees
- Decision Tree Classifier Implementation
- Decision Tree Post Pruning
- Decision Tree Regressor In depth Intuition
- Decision Tree Regressor Implementation

Part 4 - Support Vector Machines
- Support Vector Classifier Indepth Intuition
- Support Vector Machines Classifier
- Support Vector Regressor Indepth Intuition
- Support Vector Regressor Implementation
- SVM Kernels Intuition
- SVM Kernels Implementation

Part 5 - Naive Bayes
- Naive Bayes Indepth Intuition
- Variants Of Naive Bayes Algorithms
- Naive Bayes Practical Implementation

Part 6 - Ensemble Techniques & its types
- Ensemble Techniques And Bagging
- Random Forest Classifier And Regressor
- Out Of Bag Score Decision Trees
- Random Forest Practical Implementation

Part 7 – Boosting
- Boosting Technique
- Gradient Bosting Indepth Intuition
- Xgboost Classification Algorithms
- Xgboost Regresor Algorithm

Part 8 – KNN Algorithm
- KNN Classification And Regression

- Variants Of KNN
- KNN Classifier and Regr Implementation

<mark>Module 8 - Machine Learning – II – Unsupervised</mark>
Part 1- PCA
- Curse Of Dimensionality
- Geometric Intuition Behind PCA
- MAthemathical Intuition Of PCA
- PCA Practical Implementation

Part 2 – Clustering Algorithms
- K Means Indepth Intuition
- Hierarichal Clustering Intuition
- K means vs Hierarichal Clustering
- DBSCAN Clustering
- Silhouette Score Clustering
- K Means Clustering Implementation
- Hierarichal Clustering Implementation
- DBSCAN Algorithms Implementation

Part 3 – Anomaly Detection
- Anomaly Detection Isolation Fores
- DBSCAN Anomaly Detection
- Local Outlier Factor Anamoly Detection

Part 4 – Time Series
- Introduction of time series
- components of time series
- moving average
- stationary and non stationary
- acf pacf
- arima
- time sereis model building
- Time sereis EDA

<mark>Module 8 Deep Learning</mark>
Part 1 - Neural Network & Its Perception
- Intro to Deep Learning Usecases
- Neural Network, Perceptron Mathematical Explanation

Part 2 – Neural Network
- Mathematical Concepts
- Activation Functions
- Forward Back Propagation
- Implementation of ANN using Keras

- Callback Functions

Part 3 – ANN
- Regression using ANN
- Loss Function
- Batch Normalisation
- Regularization

Part 4 – Tensor Flow & Pytorch
- Regularisation in Deep Learning
- Weight Initialisation
- Optimizers
- Tensorflow
- Pytorch

Project – Credit card Fault detection

Part 5 – CNN
- Convolutional Neural Networks
- CNN Foundation 1
- CNN 2
- CNN 3 (Explainer)
- CNN 4 (LENET)
- CNN 5 (Alex Net)
- CNN 6
- CNN 7
- VGG NET
- Resnet
- Inception Net
- RCNN
- Fast RCNN
- Faster RCNN (Object Detection)
- Non Maximum Suppression

Part 7 – YOLO
- YOLO
- YOLO V2
- YOLO V3
- YOLO V4
- YOLO V4 Part 2
- YOLO V5

Part 10 – GAN
- GAN Introduction
- Training using GAN
- DC GAN
- Style GAN

- W Gan
- GAN Practical

Part 1 – NLP

- Introduction to NLP
- History of NLP
- Why NLP
- Use of NLP
- web scrapping
- Text processing
- Understanding regex

Part 2 – NLP

- String Tokenization
- Sentence Processing
- Word Embedding
- Lemmatization in text processing
- Frequency Distribution
- Annotator creation
- Word Count
- Text Normalization
- Word 2 Vec
- Co Occurance Vector

Part 3 – NLP

- Doc 2 Vec
- Text blob
- NLTK
- Genism
- RNN
- LSTM
- Bi LSTM

Module 11 Computer Vision (Open CV)

Computer Vision

- Detection of Objects
- Scene Recognition
- Image Segmentation

Project - Object racking

Project - Image Classification

Project - image to text

Module 12 - SQL

o   Basics of relational databases.

o   Basic Queries: SELECT, WHERE LIKE, DISTINCT, BETWEEN, GROUP BY, ORDER BY

o   Advanced Queries: CTE, Subqueries, Window Functions

o   Joins: Left, Right, Inner, Full

o   No need to learn database creation, indexes, triggers etc. as those things are rarely used by data scientists.

Module 13 & 14 - Power BI & Tableau

- BI tools nowadays are mainly used by BI developers, data analysts etc. Hence it is ok if you don't learn them as a data scientist. Majority of the time whenever data scientists have a need of BI dashboards they will take help of BI or data analyst teams. In small organizations however, sometimes data scientist work on building BI dashboards but in general you should not worry about learning BI tool for a data scientist career

Free Resource - http://tinyurl.com/2s4bft8k

Scan and download the App Now