EXA AI Roadmap (Based on Berkeley's Professional Certificate in Machine Learning & AI)

Master the Machine Learning and Artificial Intelligence concepts and skills from the Berkeley's Professional Certificate program, entirely for FREE! Our curated guide and resources offer a structured path to AI excellence, tailored to your individual pace.

Hi, I'm Jean

I'm the Founder and host of Exaltitude on YouTube. I've worked in tech for the past 20 years as an engineer, an engineering manager, and a team builder. I was the 19th engineer at WhatsApp and worked with Facebook as an Engineering Manager for six years after the $19B acquisition.

Throughout my career, I've mentored and coached countless Software Engineers and Managers from diverse backgrounds, noticing common questions around direction and growth: "Where am I headed, and how do I get there?" This inspired me to share my insights, helping future engineers build purposeful, successful careers.

Stay connected for updates, industry insights, and career advice on LinkedIn and YouTube. Have questions? Reach out anytime on my website!

Outline

Background Essentials - What You Need to Know

Commitment

The six-month Berkeley Professional Certificate program recommends dedicating 15–20 hours per week to watching videos and up to 8 additional hours on optional career development activities to enhance professional growth. Throughout the program, participants will complete 23 modules and a capstone project, ensuring a comprehensive learning experience.

When studying part-time, self-paced learning can take significantly longer, with the timeline influenced by several factors. Your experience level plays a key role—those with a strong foundation in math, programming, and related fields may progress more quickly.

Additionally, your dedication and the time you commit to studying will impact how fast you move through the material. The depth of your learning also matters; aiming for a deep understanding of each topic may require more time. Setting realistic expectations and being patient with yourself is crucial. Remember, the ultimate goal is to truly understand and internalize the material, not just to complete the program quickly.

Future Job Titles

This program is designed to equip you with the hands-on skills needed to launch or accelerate your career in machine learning and artificial intelligence. Upon completion, you will be prepared for a range of roles in these cutting-edge fields. Representative job titles include Data Scientist, Machine Learning Scientist, Machine Learning Engineer, and Artificial Intelligence Engineer.

Recommended Tools and Resources for the Program

We'll cover all tools as needed throughout the program, but here's a quick summary before getting started:

1. **Python**: The primary programming language used for data analysis, machine learning, and AI development.
2. **Jupyter**: An interactive environment for writing and running Python code, useful for data analysis and experimentation.
3. **Pandas**: A powerful library for data manipulation and analysis, particularly for structured data.
4. **Google Colab**: A cloud-based platform that provides a Jupyter-like environment with free access to GPUs for accelerated computing.
5. **Seaborn**: A Python visualization library based on Matplotlib, used for creating informative and attractive statistical graphics.
6. **GitHub**: A platform for version control and collaborative development, essential for sharing and tracking your projects.
7. **Codio**: A cloud-based environment for coding exercises and hands-on projects, offering an integrated platform for learning and practicing coding.
8. **Plotly**: A library for creating interactive, web-ready plots and visualizations, ideal for presenting data insights.

SECTION 1: Berkeley's Prerequisite

- An educational background in STEM fields
- Technical work experience
- Some experience with Python, R, or SQL
    - **Recommended Courses:**
        - Python Full Roadmap by Jean
        - Introduction to R by DataCamp (paid)
        - Intro to SQL by DataCamp (paid)
- Some experience with statistics and calculus
    - **Recommended Courses:**
        - Statistics and Probability by Khan Academy
        - Calculus 1 by Khan Academy
- **Optional Resource:** Full Data Science Roadmap by Jean

SECTION 2: Career Preparation and Guidance

This section is important because **learning to code and getting a job are two different skills**. Coding is about building your technical knowledge and creating projects while getting a job is about writing a strong resume, networking, and preparing for interviews. Balancing both skill sets is key to achieving your career goals.

Included in the tuition of $7900, Berkeley provides live career coaching sessions and open Q&A events, where you can gain personalized guidance. You'll also receive a bird's-eye view of the current job market landscape and its five-year trajectory, equipping you with the knowledge to navigate the future of tech careers. Additionally, the program offers valuable insights into the latest trends to enhance your chances of success in the job search process. However, there are also **free alternatives available to support your career journey!**

Resume Writing:

- The Ultimate Resume Handbook by Jean (paid): A comprehensive guide to crafting standout resumes tailored for tech roles. Also, download the free Ultimate Resume Template.
- Developer Resume with ChatGPT for ATS Success by Jean on YouTube: Learn how to use ChatGPT effectively to optimize your resume for applicant tracking systems.
- Engineering Resume Hack (from Big Tech Hiring Manager) by Jean on YouTube: Insider tips from Jean, a former hiring manager, to make your resume stand out in big tech.

Interviews:

- Cracking the Coding Interview by Gayle L. McDowell (paid book)
- Blind 75 Leet code questions by Leetcode
- Python cheat sheet by Leetcode
- DSA study guide by Leetcode
- System Design Interview Survival Guide (2024): Strategies and Tips (blog)

Join FREE Monthly LinkedIn Live Q&A with Jean:

- **What it is:** An open session where Jean answers your career-related questions live—completely free.
- **Why it matters:** Gain expert insights on navigating the tech industry, job search strategies, and career development.
- **Next session:** January 6, 2025. Sign up here. Follow Jean on LinkedIn to get updates on the newest events.

Job Market Insights:

- Tech Salaries Trends for 2025 by Jean on YouTube: Stay informed about the latest salary trends in the tech industry.
- Top Machine Learning Engineer Salary by Jean on YouTube: Explore the earning potential of ML roles in various industries.

Career Development:

- What Color Is Your Parachute? By Richard N. Bolles (Paid book)
- Zero to AI ML Engineer: Get Hired Without Experience by Jean on YouTube: A roadmap for breaking into AI/ML engineering without prior experience.
- 7 Habits That Will Make You a Better Programmer by Jean on YouTube: Simple habits to improve your coding skills and professional growth.
- 7 Mistakes that Ruin Your Career as a Junior Software Engineer by Jean on YouTube: Avoid common pitfalls that could derail your early career.

For Community:

- Join the free LinkedIn private group, Achieve Together with EXA: Connect with like-minded professionals for accountability, support, and collaboration while following this roadmap.

SECTION 3: Foundations of ML/AI

This section covers the basic concepts of machine learning, introduces key Python libraries, and helps you explore real-world applications of data science. By the end of this section, you should be comfortable with fundamental ML algorithms and data analytics tools.

Module 1: Introduction to Machine Learning

- **Description:** Introduces fundamental machine learning concepts such as supervised and unsupervised learning.
- **Course Recommendation:**
    - CS 198-126: Lecture 1 - Intro to Machine Learning on YouTube: Visit https://ml.berkeley.edu/decal/modern-cv to see more information about the course, including slides, assignments, lectures, and course information/background.

Module 2: Fundamentals of Statistics and Distribution Functions

- **Description:**
    - Learn statistical techniques such as measures of central tendency, variance, and different probability distributions.
    - Understand the importance of statistics for model evaluation and data-driven decision-making.
- **Course Recommendation:**
    - Khan Academy's Statistics and Probability: A comprehensive course that covers statistical concepts like distributions, variance, and probability, making it perfect for beginners.
    - A First Course in Probability, by Sheldon Ross, Pearson (paid resource)

Module 3: Introduction to Data Analytics

- **Description:**
    - Discover the basics of data analytics, including data collection, cleaning, and preparation.
    - Gain an overview of how analytics informs decision-making in businesses.
- **Course Recommendation:**
    - Exploratory Data Analysis in Python by DataCamp: Provides an introduction to data analytics concepts and tools for data cleaning, preparation, and visualization.
    - Python for Data Analysis by Wes McKinney (Paid book)

Module 4: Fundamentals of Data Analytics

- **Description:**
    - Build on the previous module with core data analytics concepts such as data wrangling, feature selection, and exploratory data analysis.
    - Focus on hands-on practice using Python libraries like Matplotlib and Pandas to manipulate and visualize data effectively.
- **Course Recommendation:**
    - Introduction to Data Visualization with Matplotlib by DataCamp
    - Joining Data with Pandas by DataCamp

Module 5: Practical Applications I

Here are three capstone project ideas ranging from beginner to intermediate difficulty levels. See the Capstone Project Section for more tips.

1. Beginner: Exploratory Data Analysis (EDA) on a Public Dataset

- **Goal:** Perform an in-depth exploratory data analysis (EDA) on a publicly available dataset to understand the underlying patterns, relationships, and insights.
- **Data Source:** Choose a publicly available dataset like the **Iris Dataset**, **Titanic Dataset**, or **COVID-19 dataset** (available on Kaggle or UCI Machine Learning Repository).
- **Skills Learned:**
    - Data cleaning and preprocessing (handling missing values, outliers, and duplicates).
    - Data visualization (using libraries like Matplotlib, Seaborn, and Plotly).
    - Descriptive statistics (mean, median, mode, standard deviation).
    - Correlation analysis and basic feature engineering.
    - Summarizing findings in an interactive Jupyter notebook or a dashboard.
- **Outcome:** A report or presentation showcasing key findings, trends, and visualizations from the dataset, highlighting significant insights.

2. Intermediate: Predicting House Prices Using Regression Models

- **Goal:** Develop a model to predict house prices based on various features like square footage, number of bedrooms, location, etc.
- **Data Source:** Use the Ames Housing Dataset or any other similar housing dataset.
- **Skills Learned:**
    - Data cleaning (handling missing values, encoding categorical variables, scaling).
    - Feature selection and engineering (selecting important variables, creating new features).
    - Regression models (Linear Regression, Multiple Linear Regression).
    - Model evaluation techniques (R^2, Mean Absolute Error, etc.).
    - Visualizing model performance and understanding feature importance.
- **Outcome:** A predictive model with performance metrics to evaluate its accuracy in predicting house prices, along with a comprehensive report detailing the steps involved and analysis of the results.

3. Intermediate: Customer Segmentation Using Clustering Algorithms

- **Goal:** Segment customers into distinct groups based on their purchasing behavior and demographics, helping businesses tailor marketing strategies.
- **Data Source:** Use a dataset like the Mall Customer Segmentation Dataset or any dataset with demographic and purchasing data (available on Kaggle).
- **Skills Learned:**
    - Data preprocessing (normalization, dealing with categorical variables).
    - Clustering algorithms (K-means, DBSCAN, Hierarchical Clustering).
    - Dimensionality reduction (PCA for visualizing high-dimensional data).
    - Evaluating clustering results (Silhouette Score, Elbow Method).
    - Visualizing clusters using tools like Seaborn, Plotly, and PCA.

- **Outcome:** A customer segmentation model with visualizations of the clusters and a business-driven interpretation of each segment, identifying the key characteristics of each group.

These projects will help you develop your skills in data cleaning, visualization, feature engineering, and model development, progressing from simple analysis to more complex modeling tasks.

SECTION 4: ML/AI Techniques

This section introduces essential machine-learning techniques, including clustering, regression, and regularization. You will learn how to apply algorithms like k-means and linear regression in Python, while also exploring model evaluation methods and handling overfitting.

Module 6: Clustering and Principal Component Analysis (PCA)

- **Clustering** is a method that groups similar data points together so that things in the same group are more alike than things in different groups.
    - Clustering on Scikit-Learn.
- **Principal Component Analysis (PCA)** is a technique that reduces the number of variables in a dataset by combining them into new, simpler variables while keeping most of the important information.
    - PCA on Sci-kit-learn.

Module 7: Linear and Multiple Regressions

- **Linear Regression** is a method used to predict a value by finding a relationship between that value and one or more other variables.
    - StatQuest's Linear Regression on YouTube: An accessible breakdown of linear regression concepts.

- **Multiple Regression** is a technique used to predict a value by considering multiple factors or variables that might influence it.
    - Khan Academy's Multiple Regression: A comprehensive resource explaining multiple regression and its applications.

Module 8: Feature Engineering and Overfitting

- **Feature Engineering** is the process of creating new input features or modifying existing ones to improve the performance of a machine-learning model.
    - Kaggle Micro-Course: Feature Engineering
- **Overfitting** happens when a model learns the details of the training data too well, including noise or errors, making it perform poorly on new, unseen data.
    - What is Overfitting (blog)

Module 9: Model Selection and Regularization

- **Model Selection:**
    - Learn techniques to evaluate and choose the best-performing model for a given dataset, including cross-validation.
    - Model selection and evaluation on Scikit Learn: Covers model evaluation and selection.
- **Regularization** is a technique used to prevent overfitting by adding a penalty to the model, making it simpler and less likely to fit noise in the data.
    - Google's Crashcourse on Overfitting: L2 regularization: Provides insights into real-world applications of regularized models.

Module 10: Time Series Analysis and Forecasting

- **Time Series Analysis** is the process of analyzing data points collected or recorded at specific time intervals to identify trends, patterns, and make predictions about the future.
- **Forecasting** is the process of predicting future events or trends based on past data.

- **Resource:** Kaggle Micro-Course on Time Series and Forecasting: Focuses on time series data manipulation and visualization.

Module 11: Practical Application II

Here are two capstone project ideas incorporating Clustering, PCA, Regression, Feature Engineering, Regularization, and Time Series Analysis. See the Capstone Project Section for more tips.

1. Customer Segmentation and Behavior Analysis

- **Goal:** Segment customers into distinct groups based on purchasing behavior and demographic features, and forecast their future spending.
- **Techniques Used:**
    - **Clustering & PCA:** Use clustering (e.g., K-means) to group customers with similar behaviors. Apply PCA for dimensionality reduction to visualize customer data in fewer dimensions while retaining key features.
    - **Feature Engineering & Overfitting:** Engineer features such as customer tenure, average purchase value, and frequency of purchases. Use regularization (Lasso or Ridge) to avoid overfitting in your clustering models and regression analysis.
    - **Linear and Multiple Regressions:** Implement linear regression to predict future spending based on demographic and behavioral features. Use multiple regression to predict more complex relationships between customer features and spending habits.
    - **Model Selection & Regularization:** Apply cross-validation to select the best model and prevent overfitting using regularization techniques such as L1 (Lasso) or L2 (Ridge) regularization.

2. Sales Forecasting for Retail Business

- **Goal:** Predict future sales for various products in different regions, accounting for seasonality and external factors.

- **Techniques Used:**
    - **Time Series Analysis & Forecasting:** Use time series techniques (e.g., ARIMA or Prophet) to model and forecast sales data, incorporating seasonality and holidays.
    - **Clustering & PCA:** Apply clustering algorithms to group stores or products with similar sales patterns. Use PCA to reduce dimensionality in product features while maintaining key sales drivers.
    - **Feature Engineering & Overfitting:** Engineer features such as promotional periods, holidays, and store location. Regularize your models using Lasso or Ridge to avoid overfitting to noisy historical data.
    - **Linear and Multiple Regressions:** Use linear regression to forecast sales for each store, considering factors like pricing, promotions, and weather conditions. Use multiple regression models to understand the combined effect of various factors on sales.

These projects incorporate a variety of advanced techniques and can be tailored to real-world business needs, making them great choices for a capstone project. See the Capstone Project Section for more tips.

Module 12: Classification and k-Nearest Neighbors (kNN)

- **Classification** is the process of sorting data into categories based on certain features or characteristics.
    - Classification in Machine Learning: An Introduction (blog)
- **k-Nearest Neighbors (kNN)** is a method that classifies data based on the closest data points to it.
    - StatQuest kNN on YouTube: A clear, engaging introduction to kNN algorithms.

Module 13: Logistic Regression

- **Logistic Regression** is a technique used to predict the probability of an outcome based on input data, often used for binary classification (like yes or no).

- Understanding Logistic Regression in Python (blog)

Module 14: Decision Trees

- **Decision Trees** are a model that splits data into branches to make decisions or predictions based on features.
- **Course Recommendation:** Decision Trees by StatQuest on YouTube

Module 15: Gradient Descent and Optimization

- **Gradient Descent** is an optimization method used to find the best model by minimizing errors through repeated adjustments.
    - Gradient Descent Visualizations by 3Blue1Brown on YouTube
- **Optimization** is the process of adjusting a model to find the best possible solution by reducing errors.
    - Introduction to Optimization in Python on DataCamp

Module 16: Classifying Nonlinear Features

- **Classifying Nonlinear Features** involves using techniques that can handle data with complex, non-straightforward patterns to make accurate predictions.
- **Course Recommendation:** Non Linear Classification - Ep.4 (Deep Learning Fundamentals)on YouTube: Covers nonlinear classification using neural networks and support vector machines.

Module 17: Practical Application III

Here are three examples of capstone projects that utilize classification techniques, KNN, Logistic Regression, Decision Trees, Gradient Descent, and Optimization. See the Capstone Project Section for more tips.

1. Customer Churn Prediction using Classification Algorithms:

- **Goal:** Predict customer churn for a subscription-based service (e.g., telecom, streaming).

- **Techniques Used:**
    - **Classification:** Use classification algorithms to predict whether a customer will churn or not.
    - **KNN:** Apply K-Nearest Neighbors to classify customer churn based on similar user behaviors.
    - **Logistic Regression:** Implement logistic regression to model the probability of churn based on customer features.
    - **Decision Tree:** Build decision tree models to identify important factors influencing churn.
    - **Optimization:** Use gradient descent to optimize model parameters.

2. Handwritten Digit Recognition:

- **Goal:** Classify images of handwritten digits from the MNIST dataset.
- **Techniques Used:**
    - **Classification:** Build a model to classify digits from images.
    - **KNN:** Use KNN for pattern recognition and classification based on pixel intensity features.
    - **Logistic Regression:** Apply logistic regression to map pixel values to digit probabilities.
    - **Gradient Descent & Optimization:** Train the model using gradient descent to minimize classification error.

3. Credit Card Fraud Detection:

- **Goal:** Detect fraudulent transactions using historical transaction data.
- **Techniques Used:**
    - **Classification:** Classify transactions as either fraudulent or legitimate.
    - **Logistic Regression:** Use logistic regression to model the probability of fraud.
    - **Decision Tree:** Use decision trees to understand which transaction attributes lead to fraud.

- **Gradient Descent & Optimization:** Apply gradient descent for model training to minimize loss.

SECTION 5: Advanced Machine Learning Techniques

This section covers advanced machine learning and AI techniques, focusing on real-world applications in various industries. You will explore key concepts like natural language processing, recommendation systems, ensemble techniques, and generative AI, building expertise in solving complex ML/AI problems.

Module 18: Natural Language Processing

- **Description:** Learn techniques for processing and analyzing textual data, focusing on tasks like sentiment analysis, text classification, and named entity recognition (NER).
- **Course Recommendation:**
  - Kaggle's Natural Language Processing with Python: Practical exercises and tutorials for building NLP models in Python.
  - Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems by Sowmya Vajjala (Paid book)

Module 19: Recommendation Systems

- **Description:** Explore collaborative filtering, content-based filtering, and hybrid models used in recommendation systems for personalizing user experiences.
- **Course Recommendation:** Recommender Systems Specialization on Coursera: In-depth exploration of various recommendation algorithms and their applications.

Module 20: Ensemble Techniques

- **Description:** Learn how ensemble methods, such as Random Forests and Gradient Boosting, improve model performance by combining multiple weak learners into a stronger one.
- **Course Recommendation:** Kaggle's Ensemble Learning Technique Tutorial: Practical tutorials for implementing ensemble techniques in machine learning projects.

Module 21: Deep Neural Networks I

- **Description:** Understand the foundational concepts of deep neural networks (DNNs), including feedforward networks, backpropagation, and training algorithms.
- **Course Recommendation:** Fast.ai's Practical Deep Learning for Coders: Hands-on course focused on building deep learning models.

Module 22: Neural Networks II

- **Description:** Explore advanced deep learning topics, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), and their applications in image and sequence data.
- **Course Recommendation:** Coursera's Convolutional Neural Networks by Andrew Ng: Focuses on CNNs for image recognition tasks.

Module 23: Introduction to Generative AI

- **Description:** Learn the basics of generative AI, focusing on generative adversarial networks (GANs) and autoencoders, and explore how these techniques can create new data and solve real-world problems.
- **Course Recommendation:** Coursera's Generative Adversarial Networks (GANs) Specialization: A hands-on course on building and training GANs.

Module 24: Capstone Project

Here are three capstone project ideas that incorporate Natural Language Processing (NLP), Recommendation Systems, Ensemble Techniques, Deep Neural Networks (DNN), and Generative AI:

1. Personalized Content Recommendation System Using NLP and Deep Neural Networks

- **Goal:** Build a recommendation system that suggests articles, blogs, or books based on a user's reading history and preferences.
- **Techniques Used:**
    - **NLP:** Use NLP techniques to analyze and extract features from text-based content (e.g., TF-IDF, word embeddings like Word2Vec or BERT) to understand the semantic meaning of articles.
    - **Recommendation Systems:** Implement collaborative filtering and content-based filtering to recommend articles based on user behavior and preferences.
    - **Ensemble Techniques:** Use ensemble methods such as Random Forest or Gradient Boosting to combine multiple models for more accurate recommendations.
    - **Deep Neural Networks:** Build a deep neural network model to learn complex patterns in user behavior and content features to improve the recommendation system's performance.
    - **Generative AI:** Use a generative model like GPT to generate personalized content summaries or new article ideas based on the user's reading pattern.

2. Sentiment Analysis and Product Review Summarization Using NLP and Generative AI

- **Goal:** Develop a system that performs sentiment analysis on product reviews and automatically generates concise summaries for products.
- **Techniques Used:**

- **NLP:** Use NLP techniques like tokenization, named entity recognition (NER), and sentiment classification models to analyze customer reviews and extract key sentiments.
- **Deep Neural Networks:** Use pre-trained models such as BERT or GPT for sentiment classification, and fine-tune them for the specific task of sentiment analysis on product reviews.
- **Generative AI:** Implement generative models (like GPT) to automatically generate short, meaningful summaries of product reviews while preserving the sentiment and key features.
- **Ensemble Techniques:** Apply ensemble techniques like stacking to combine multiple models (e.g., logistic regression, decision trees) for sentiment analysis to improve accuracy.
- **Recommendation Systems:** Enhance the system by recommending products based on users' sentiment preferences, i.e., recommending highly rated products with positive sentiments.

3. AI-Driven Music Playlist Generation and Personalization

- **Goal:** Build an AI system that generates personalized music playlists based on a user's listening history and mood.
- **Techniques Used:**
  - **NLP:** Use NLP to analyze song lyrics and extract features such as sentiment, theme, or genre for better personalization of music recommendations.
  - **Recommendation Systems:** Implement a hybrid recommendation system that combines user-based collaborative filtering and content-based filtering to suggest songs based on a user's music preferences and behavior.
  - **Deep Neural Networks:** Develop a deep learning model (such as a neural collaborative filtering model) to learn complex patterns in user preferences and music features.

- **Ensemble Techniques:** Use ensemble methods like XGBoost or LightGBM to improve recommendation accuracy by combining different models (e.g., content-based, collaborative, and hybrid).
- **Generative AI:** Leverage a generative AI model to create new songs or remixes based on the user's preferred music style, genre, and mood, enhancing the playlist experience.

These projects offer a diverse set of challenges that integrate cutting-edge techniques in NLP, deep learning, recommendation systems, and generative AI, making them ideal for a capstone project in AI.

SECTION 6: Capstone project

Goal

Aim to get your projects to publishable quality for submission to conferences or journals. For inspiration, review recent machine learning research papers from major conferences like ICML and NeurIPS, or look at Stanford's past class projects for ideas.

Research Papers

For further studies and in-depth knowledge, consider reading research papers to stay updated with the latest advancements in the field. Some of the most reputable databases to find cutting-edge research include:

- Arxiv Sanity: allows you to browse state-of-the-art, trending research in various AI and machine learning domains.
- Browse State-of-the-Art Trending Research: explore the latest, most impactful papers and developments in AI and machine learning.
- Deep Learning Monitor: offers comprehensive resources and papers focused specifically on deep learning.

- Distilled AI List of research papers since 2010: provides a curated collection of key papers that have shaped AI research over the years, offering valuable insights into the evolution of the field.

Exploring these platforms will deepen your understanding and help you stay ahead in the ever-evolving world of AI and machine learning.

For tips on how to read research papers, watch How to actually learn AI/ML: Reading Research Papers by Jean on YouTube.

Berkeley's Project Ideas

These are some ideas for coding exercises from Berkeley designed to help you practice and build composite skills necessary for assignments and portfolio projects:

1. **Train a Decision Tree Model with Desired Hyperparameters Using Scikit-learn:** Implement a decision tree model using Scikit-learn. You will learn how to tune hyperparameters to optimize the model's performance and make predictions based on input features.
2. **Plot Decision Boundaries Using Logistic Regression:** Apply logistic regression to a classification problem and visualize decision boundaries. This exercise will help you understand the distinction between different types of classifiers and their ability to separate classes in feature space.
3. **Construct a Model Using Classical Time Series Decomposition:** This activity will teach you how to decompose a time series into its components (trend, seasonality, and noise). You will use classical techniques to build a model that can forecast future values.
4. **Import and Clean Messy Data from Real-World Datasets: P**ractice working with raw, unstructured datasets by cleaning and preparing them for analysis. This exercise includes handling missing values, correcting data types, and dealing with outliers to make the data ready for modeling.
5. **Create Histograms and Data Visualizations in Python:** Use Python libraries like Matplotlib and Seaborn to create histograms and various other data

visualizations. This will help you communicate data insights effectively through visual representations.

6. **Perform Computations Between DataFrames Using Set Index and Reset Index:** Master pandas DataFrame operations, particularly using the `set_index` and `reset_index` methods to perform computations and manipulate data efficiently for analysis.
7. **Perform String Manipulation in Pandas:** Clean and transform textual data using pandas. String manipulation is a critical skill when dealing with real-world data, and you'll learn how to handle operations like string matching, substitution, and extraction.
8. **Apply Singular Value Decomposition (SVD) to a Specific Dataset:** Singular Value Decomposition (SVD) is a powerful matrix factorization technique used for dimensionality reduction. In this exercise, you will apply SVD to reduce the dimensions of a dataset and analyze the resulting components.

Utilizing ChatGPT

ChatGPT can be a valuable tool for creating detailed project plans.

Sample Prompt: Tell me how I can approach building this project in *[insert project description.]* Give me practical hands-on resources and a step-by-step guide.

Additional Tips:

- Consistency is vital: Dedicate a specific time each day for studying.
- Take breaks: Avoid burnout by taking short breaks.
- Join online communities: Connect with other learners for support and collaboration.
- Build projects: Apply your knowledge by creating small projects.
- Stay motivated: Set achievable goals and celebrate your progress.

Remember, the actual time will vary depending on your learning pace and prior knowledge.

Good luck!