AI & Machine Learning SkillJourney

Machine Learning Essentials Boot Camp / Part 1: Preparing Your Data

Empower your machine learning models with the best in data preprocessing and analysis techniques

Snapshot

- **Course: Machine Learning Essentials Boot Camp / Part 1: Preparing Your Data (TTML5510)**
- **Duration**: 3 days (or 5 half days)
- **Audience & Skill Level:** This is an intermediate-level program, designed to prepare attendees for a deeper dive into next-level, heavy hands-on machine learning courses and workshops. Attendees should have practical, hands-on experience working with Python for Data Science, pandas and numpy.
- **Format / Hands-on:** This course combines engaging instructor-led presentations and practical demonstrations with hands-on programming exercises, challenge labs, use case exploration and engaging group activities. Student machines are required.
- Flexible Delivery Options: This course can be delivered for your team or organization online-live (virtual), onsite in-person, self-paced or across our immersive blended learning experience platform (LXP).
- **Public Schedule**: This course is currently available on our Public Open Enrollment Schedule.
- **Customizable**: Were flexible! This course agenda, topics, labs, hours and delivery modalities can be adjusted to target your specific training skills objectives, tools and learning goals. Please ask for details.

Overview

In the world of machine learning, the quality of input data is critical. Machine learning models that use bad data input produce inaccurate and unreliable results, undermining their effectiveness and trustworthiness. Our **Machine Learning Essentials Boot Camp: Preparing Your Data** is a three-day hands-on skills immersion course geared for students who need to how to effectively prepare and optimize data for use in machine learning models, ensuring they produce accurate, useful and insightful predictions.

Throughout the course, guided by our expert instructor, you'll engage in workshop-style practical labs that will provide you with the real-world skills and hands-on experience needed to manage, prep and clean your data for successful machine learning model applications.

You'll learn how to translate diverse data into an analytically-friendly format, ensuring compatibility with machine learning algorithms. You'll learn how to scale and normalize data, ensuring consistent data representation, crucial for accurate model training and predictions. You'll navigate the intricacies of data transformation and refinement, and learn how to translate diverse datasets into formats friendly to machine learning algorithms. You'll also explore feature selection and dimensionality reduction, striking the balance between data richness and computational efficiency. You'll also grasp how to safeguard your data's journey with robust pipelines and preventive measures against data leakage, cementing the trustworthiness of your real-world model deployments. Lastly, you'll explore the complete lifecycle of a machine learning project, from data preparation to model deployment, you're equipped to oversee and implement comprehensive data-driven solutions.

By the end of this immersive boot camp, you'll be fully-equipped with a comprehensive skillset that not only enhances the predictive power of your models but also sets the foundation for innovative, data-driven solutions. You'll be ready to advance in your Machine Learning journey, leveraging your newly acquired skills towards model proficiency.

Objectives

Throughout the course you will explore:
- Data Encoding: Dive into data encoding to seamlessly translate diverse information into a machine-friendly format.
- Data Manipulation Mastery: You'll get comfortable with encoding, scaling, and normalizing data. By the end of the course, the curse of dimensionality will no longer be a challenge.

- Quality Analysis Confidence: Learn how to identify and remove duplicates, handle null values, manage outliers, and work with dates in your data. You'll be a pro at maintaining clean datasets.
- Feature Analysis Wizardry: Discover how to identify unused columns, detect low variance ones, and understand multicollinearity. By the end of the workshop, feature selection will feel like second nature.
- Pipeline Proficiency: Gain a deep understanding of the critical role of pipelines in machine learning and develop the skills to create and implement your own data preprocessing pipelines.
- Machine Learning Basics: Get introduced to the fundamentals of machine learning, understand k-fold cross-validation, master the art of partitioning data, and learn how to prevent data leakage. You'll be set to step confidently into the world of machine learning.

Audience

This course is geared for data scientists and business professionals seeking to leverage data insights in decision-making. It's also ideal for software developers wanting to diversify their skills into the exciting field of machine learning. Whether you're a student eager to jumpstart your career or an experienced professional looking to enhance your data-driven strategies, our hands-on workshop offers a valuable learning experience to transform you into a confident data handler and problem-solver.

Pre-Requisites

This is an intermediate-level program, designed to prepare attendees for a deeper dive into next-level, heavy hands-on machine learning courses and workshops. Attendees should have practical, hands-on experience working with Python for Data Science, pandas and numpy.

Take Before: Students should have incoming practical skills aligned with those in the course(s) below, or should have attended the following course(s) as a pre-requisite:
- TTPS4873   Fast Track to Python for Data Science
- TTPS4874   Applied Python for Data Science

Take After/ Follow On: The following course(s) might serve as good next steps in your learning journey.
- TTML5511   Machine Learning Essentials Boot Camp (Part 2): Machine Learning Deep Dive (3 days)
- TTML5503   Introduction to AI, AI Programming and Machine Learning (3 days)

Next Steps / Follow-on Courses: We offer a wide variety of related or follow-on courses and learning paths for Machine Learning, Data Science, Advanced Python, Artificial Intelligence, AI for Business, Generative AI, GPT, AutoGPT, AGI, Applied AI, Azure OpenAI, Google BARD, AI for developers, testers, data analytics, deep learning, programming, intelligent automation and many other related topics. Please see our catalog for the current **AI & Machine Learning Courses, Learning Journeys & Skills Roadmaps**, list courses and programs.

Setup Made Simple! Learning Experience Platform (LXP)

All applicable course software, digital courseware files or course notes, labs, data sets and solutions, live coaching support channels and rich extended learning and post training resources are provided for you in our "easy access, no install required" online **Learning Experience Platform (LXP),** remote lab and content environment.

Program Agenda

Please note that this list of topics is based on our standard course offering, evolved from typical industry uses and trends. We'll work with you to tune this course and level of coverage to target the skills you need most. Topics, agenda and labs are subject to change, and may adjust during live delivery based on audience skill level, interests and participation.

1. **Getting Started with Data**
- Explore the role and importance of data in machine learning.

- Encoding data: Transform raw data into a format suitable for analytics.
- Dealing with the curse of dimensionality: Navigate

high-dimensional spaces effectively.
- Scaling and normalizing data: Standardize data for consistent analysis.
- Hands-on Activity / Lab

2. **Structural Analysis**
- Delve into the intricate patterns that define data.
- Importing libraries: Equip yourself with the right tools for data manipulation.
- Importing data: Initiate the first steps of data-driven exploration.
- Conducting basic data investigation: Peek into the essence of your dataset.
- Utilizing relevant tools for data structure analysis: Get acquainted with state-of-the-art tools to dissect data structure.
- Hands-on Activity / Lab

3. **Quality Analysis**
- Refine data sets by spotting and fixing errors.
- Identifying and removing duplicates: Ensure uniqueness in your dataset.
- Handling null values and missing data: Fill the gaps in your data with precision.
- Detecting and managing outliers: Understand and manage extreme data points.
- Working with dates in data: Harness the power of time-series data.
- Hands-on Activity / Lab

4. **Exploratory Data Analysis**
- Dive deep into data to extract meaningful insights.
- Conducting univariate analysis: Analyze one variable at a time.
- Conducting bivariate analysis: Discover relationships between two variables.
- Conducting multivariate analysis: Understand complex data interactions.
- Using pivot tables for data analysis: Summarize data visually and numerically.
- Understanding correlation: Measure linear relationships between variables.
- Understanding mutual information: Gauge dependency between variables.
- Hands-on Activity / Lab

5. **Data Features**
- Pinpoint the most impactful data components.
- Identifying and dropping unused columns: Streamline data for efficiency.

- Detecting and handling low variance or no variance columns: Maintain data variability.
- Understanding multicollinearity (VIF): Ensure independent predictor variables.

6. **Feature Selection**
- Prioritize the most relevant data features for robust models.
- Using wrappers (RFE, Forward, Backward selection): Implement dynamic feature selection.
- Using filters (Statistical tests): Opt for features based on statistical relevance.
- Using embedded methods: Integrate feature selection into algorithm functionality.
- Understanding unsupervised feature selection methods: Navigate feature selection without target variables.
- Hands-on Activity / Lab

7. **Feature Importance**
- Gauge the significance of different data features in prediction.
- Understanding dimensionality reduction: Simplify data without losing information.
- Using Principal Component Analysis (PCA): Transform data to highlight variance.
- Using Linear Discriminant Analysis (LDA): Optimize class separability.
- Hands-on Activity / Lab

8. **Encoding, Scaling, and Skewness**
- Tailor data formats for better compatibility with machine learning algorithms.
- Encoding categorical variables: Convert categories into numerical values.
- Scaling numerical variables: Maintain consistency in data magnitude.
- Detecting and correcting skewness in data: Normalize data distributions.
- Hands-on Activity / Lab

9. **Pipelines**
- Streamline machine learning workflows with seamless data transitions.
- Understanding the role of pipelines in machine learning: Appreciate the significance of efficient workflows.
- Creating and implementing data preprocessing pipelines: Process data in a structured manner.
- Using pipelines for efficient cross-validation and

hyperparameter tuning: Optimize model parameters with ease.
- Hands-on Activity / Lab

10. Introduction to Machine Learning
- Lay the groundwork for next-level machine learning practices.
- Understanding k-fold cross-validation: Assess model performance effectively.
- Using resampling techniques: Balance dataset disparities.

- Dividing data into training and test sets: Create a structured environment for model training and evaluation.
- Identifying and preventing data leakage: Maintain the integrity of your datasets.
- Understanding the basic types and applications of machine learning models
- **Capstone Project:** Develop an end-to-end machine learning model: Apply the course skills to develop a complete data-driven projects.

For More Information

For more information about our training services (instructor-led, self-paced or blended), collaborative coaching services, robust Learning Experience Platform (LXP), Career Experiences, public course schedule, partner programs, courseware licensing options or to see our complete list of course offerings, solutions and special offers, please visit us at **www.triveratech.com**, email **Info@triveratech.com** or call us toll free at **844-475-4559.** Our pricing and services are always satisfaction guaranteed.

TRIVERA TECHNOLOGIES ● **Collaborative IT Training, Coaching & Skills Development Solutions**
www.triveratech.com ● **toll free +1-844-475-4559** ● **Info@triveratech.com** ● **Twitter TriveraTech**

ONSITE, ONLINE & BLENDED TRAINING SOLUTIONS ● PUBLIC / OPEN ENROLLMENT COURSES
LEARNING EXPERIENCE PLATFORM (LXP) ● COACHING / MENTORING ● ASSESSMENTS ● CONTENT LICENSING & DEVELOPMENT
LEARNING PLAN DEVELOPMENT ● SKILLS IMMERSION PROGRAMS / RESKILLING / NEW HIRE / BOOT CAMPS
PARTNER & RESELLER PROGRAMS ● CORPORATE TRAINING MANAGEMENT ● VENDOR MANAGEMENT SERVICES