THE ODSC GUIDE TO MACHINE LEARNING

39 ODSC resources to guide you from machine learning beginner to expert.

INTRODUCTION

Machine learning is arguably the most popular starting point for anyone looking to get into data science, and where many organizations start to implement AI into their practice. With AI becoming commonplace in much of the world, even the term “machine learning” itself is starting to become common knowledge.

It’s difficult to say where one should start their machine learning journey. What coding language do you want to use? What tools, platforms, and libraries to work with? What problems do you need to look out for and to avoid?

What about those who already have machine learning experience? How up-to-date are your core mathematical skills? Have you tried any new frameworks out, or are you still relying on the ones you used five years ago?

These written tutorials and videos from past ODSC conferences will hopefully be able to address those questions.

Covering from the bare basics with R and Python to avoiding the black box problem, these ODSC resources will help those new to data science get started with machine learning, and already-established pros revisit their toolkit.

TOP 24 BLOGS

Much of the content on **OpenDataScience.com** revolves around machine learning modeling. Whether you're just starting out in your machine learning adventure, or you're a seasoned expert looking to try something new, there's an article here for you.

TOP BLOGS

Best Machine Learning Research of 2020

Daniel Gutierrez

Between new deep learning and NLP tools, COVID tracking, automation, and more, this is the standout machine learning research from 2020.

How Bayesian Machine Learning Works

Stefan Jansen

You keep hearing about it, now you want to see how Bayesian machine learning works. Get those answers + examples here!

Understanding the Mechanism and Types of Recurrent Neural Networks

Yuxi (Hayden) Liu

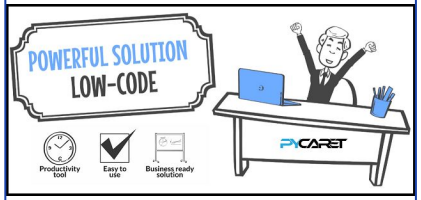
In this article, we will talk about the mechanism and types of the model used for modern sequence learning – Recurrent Neural Networks

The Ultimate Free Machine Learning Development Stack

Nick Acosta

This free machine learning development stack is all you need to create end-to-end ML pipelines with ease.

TOP BLOGS



Introduction to PyCaret

Daniel Gutierrez

Learn more about the ML library PyCaret, which can be used to perform complex machine learning tasks with only a few lines of code.

TensorFlow

Why TensorFlow Will Stand Out on Your Resume in 2021

Alex Landa

Why TensorFlow you ask? Well, there are plenty of reasons why you should be using TensorFlow for your machine and deep learning needs.

The 5 Skills You Need to Start Machine Learning

ODSC Team

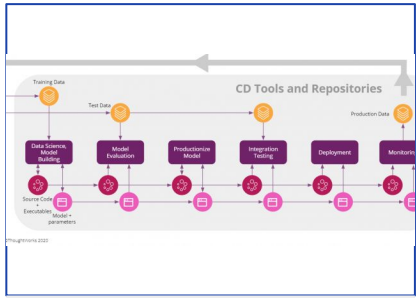
Like any new skill set, it's hard to jump right in. But if you want to know how to start machine learning, then start with these 5 skills.

A Deep Dive into H2O's AutoML

Parul Pandey

AutoML is fundamentally changing the face of ML-based solutions today by enabling people from diverse backgrounds to use machine learning models to address complex scenarios.

TOP BLOGS



Continuous Delivery for Machine Learning

Multiple

Learn more about CD4ML, which is the discipline of bringing continuous delivery principles and practices to machine learning applications.

Why You Should be Using Jupyter Notebooks

Daniel Gutierrez

As many data science professionals begin to work remotely, it's a good time to consider using Jupyter Notebooks for your machine learning projects.

How You Can Use Federated Learning for Security & Privacy

Daniel Gutierrez

New machine learning methodologies like federated learning have been developed to address concerns in privacy and security.

Unsupervised Learning with k-means Clustering With Large Datasets

Denis Rothman

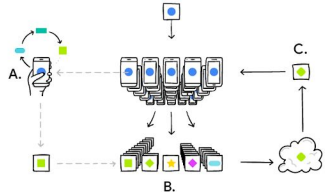
In this article, we will explore how to implement k-means clustering with dataset volumes that exceed the capacity of the given algorithm.

TOP BLOGS

Explore Fundamental Concepts of Reinforcement Learning

Giuseppe Bonaccorso

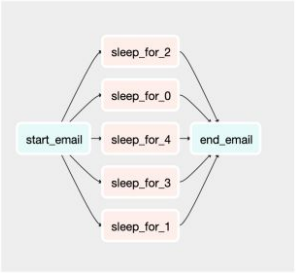
Let's take a look at some important and fundamental concepts of reinforcement learning to get you up to speed on this popular topic.



What is Federated Learning?

Daniel Gutierrez

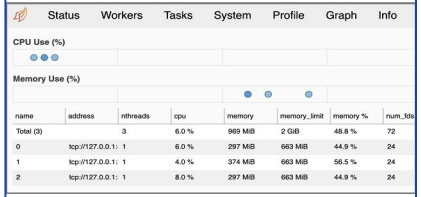
In this article, we'll explore federated learning in terms of its beginnings, benefits, challenges, as well as some recent advances.



Introduction to Apache Airflow

Tomasz Urbaszek

Get a little help with your machine learning workflow and check out this introduction to Apache Airflow and see what it can do to help.

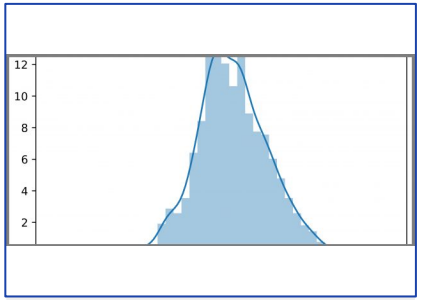


Scaling LightGBM with Dask

James Lamb

In this article, you'll learn how to use Python and Dask to take advantage of distributed LightGBM training.

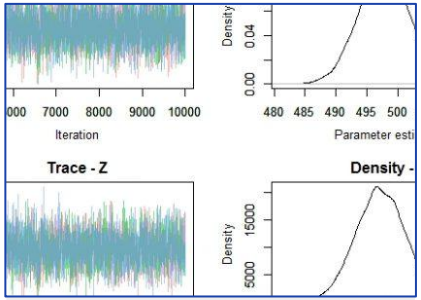
TOP BLOGS



Transforming Skewed Data for Machine Learning

Nathaniel Jermain

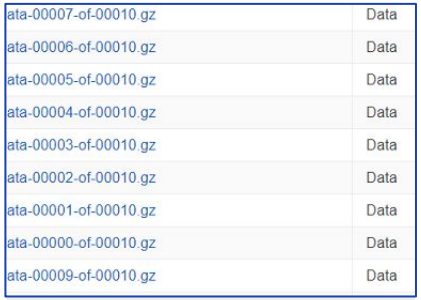
Skewed data is common in data science; skew is the degree of distortion from a normal distribution. So, let's learn about transforming skewed data.



Building Your First Bayesian Model in R

Nathaniel Jermain

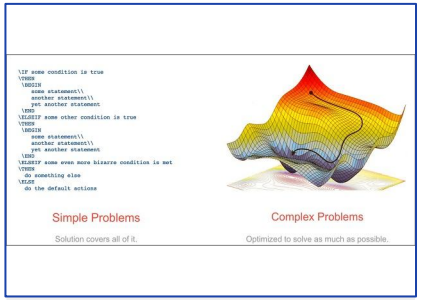
This step-by-step guide will help you use R to build your first Bayesian model, which are models that offer a method for making probabilistic predictions about the state of the world.



25 Excellent Machine Learning Open Datasets

Elizabeth Wallace

Here are our top 25 picks for open source machine learning datasets. Each one offers clean data with neat columns and rows so that your training sets run more smoothly. Let's take a look.



Dealing with the Incompleteness of Machine Learning

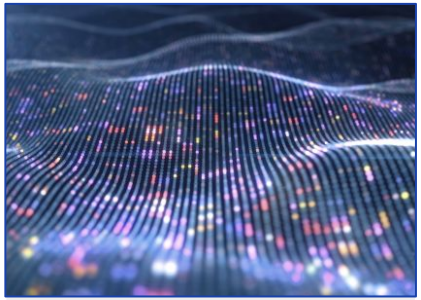
Serg Masís

By explaining a model's decisions, we can cover gaps in our understanding of the problem - it's incompleteness.

TOP BLOGS

Auto-Sklearn: AutoML in Python
Matthias Feurer, Katharina Eggensperger, and Frank Hutter

In this post, you'll learn how to replace a manually designed scikit-learn pipeline with an Auto-sklearn estimator.

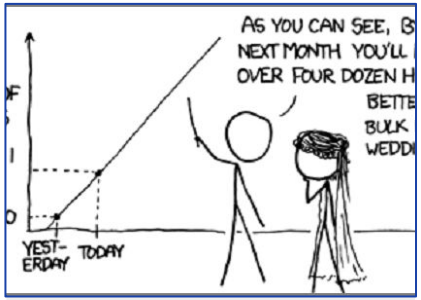


Federated Learning 101 with FEDn
Daniel Zakrisson

Here's a short and easy-to-follow tutorial to set up your first federated learning project with FEDn.

ModelOps – AI Model Operationalization for the Enterprise
Daniel Gutierrez

ModelOps represents a holistic approach for quickly and iteratively advancing models through the machine learning life cycle so they are deployed more rapidly and deliver desired business value.



Prophet's Forecasting Crystal Ball
Marta Markiewicz

Facebook's Prophet is one of the most-liked forecasting approaches nowadays. Let's take a look under the hood and see how you can use it.

TOP 15 VIDEOS

ODSC speakers cover a wide range of topics that are pivotal for any practicing data scientist or researcher.
These are 15 talks from past conferences that were highly-rated by attendees who were also looking to explore machine learning. Click on this link to see the full playlist with all of these sessions.

WATCH VIDEOS
HERE

TOP VIDEOS

Machine Learning in R Part I: Penalized Regression and Boosted Trees

Jared Lander

Linear regression is the foundation of supervised learning, though it has its limits. During this workshop we extend regression using penalization for automated variable selection and increased flexibility. We then introduce trees, and in particular boosted trees, via xgboost to get incredibly powerful predictions. We will go over some of the theory and also practical considerations such as hyperparameters.

Machine Learning in R Part II: Using workflows to build an ML optimization pipeline

Jared Lander

Modern machine learning is mostly brute forcing through a multitude of hyperparameters. We look at new tools for building and tuning models in R, some so new they are only available on GitHub. We use xgboost and glmnet models (learned in Part I) as motivation for learning how to split data and conduct cross-validation with rsample, perform feature engineering with recipes, build model specifications with parsnip, tune over hyperparameters with dials and tune, evaluate performance with yardstick and put it all together with workflows.

Solving the Data Scientist's Dilemma: the Cold-Start Problem with 10+ Machine Learning Examples

Dr. Kirk Borne

In this talk from East 2020, Kirk Borne presents 10+ machine learning examples and suggested solutions of cold-start problems (i.e., that move from a bad initial random guess to a good, perhaps optimal, solution), covering a variety of different algorithms and applications, focused primarily on unsupervised learning, but with some supervised learning examples also.

SLIDES

TOP VIDEOS

Removing Unfair Bias in Machine Learning

Margriet Groenendijk, PhD

In this workshop, you will learn the debiasing techniques that can be implemented by using the open-source toolkit AI Fairness 360, which is an extensible, open-source toolkit for measuring, understanding, and removing AI bias.

SLIDES

SLIDES

Explainable ML: Application of Different Approaches

Violeta Misheva, PhD

In the talk, Violeta will briefly go over a unique business problem and the approach her team took to solve it, as well as explain what Shapley value is, and how it can be used in many applications. With the increasing popularity of machine learning models, and the importance of transparent and explainable models in certain domains, explainability will become more and more important.

SLIDES

SLIDES

Echo State Networks for Time-Series Data

Teal Guidici, PhD

In this session, participants will be introduced to Echo State Networks (a type of recurrent neural network) including theory, key parameters in implementation and practical considerations.

SLIDES

TOP VIDEOS

Introduction to Scikit-learn: Machine learning in Python

Thomas Fan

We will start this training by learning about scikit-learn's API for supervised machine learning. scikit-learn's API mainly consists of three methods: fit, to build models, predict, to make predictions from models, and transform, to change the representation of the input data.

SLIDES

SLIDES

End to End Modeling & Machine Learning

Jordan Bakerman, PhD | Ari Zitin

In this workshop, you will load data into memory, prepare input variables for modeling and build complex analytics pipelines to demonstrate powerful machine learning models. Need to integrate open source models? No problem. We'll show you how you to do that and deploy any model.

Missing Data in Supervised Machine Learning

Andras Zsom, PhD

This workshop reviews the three types of missing data (missing completely at random, missing at random, missing not at random) and a couple of simple but often misleading ways to impute. It then describes three advanced methods for handling missing data: multiple imputation, the reduced-feature (aka pattern submodel) approach, and XGBoost.

SLIDES

TOP VIDEOS

Data Science Best Practices: Continuous Delivery for Machine Learning

Christoph Windheuser, PhD | David Johnston, PhD | Eric Nagler

In this workshop, we show how to maintain data science productivity as well as collaborate effectively and deliver value continuously and seamlessly. Participants will learn how to utilize new patterns of repeatable continuous model development to collaborate effectively and deliver value continuously and seamlessly in industrial data science projects.

SLIDES

Testing Production Machine Learning Systems

Josh Tobin, PhD

In this talk, we argue for the importance of testing in ML, give an overview of the types of testing available to ML practitioners, and make recommendations about how you can start to incorporate more robust testing into your ML projects.

Machine Learning for Biology and Medicine

Sriram Sankararaman, PhD

Biology and medicine are deluged with data so that techniques from machine learning and statistics will increasingly play a key role in extracting insights from the vast quantities of data being generated. This session provides an overview of the modeling and inferential challenges that arise in these domains.

TOP VIDEOS

Rule Induction and Reasoning in Knowledge Graphs

Daria Stepanova, PhD

This tutorial presents state-of-the-art rule induction methods, recent advances, research opportunities as well as open challenges along this avenue.

SLIDES

Bayesian Modeling without the Math

Thomas Wiecki, PhD

Bayesian modeling is an extremely powerful tool in solving data science problems across different domains. And while user-friendly modeling packages like PyMC3 exist, understanding the underlying concepts still provides a challenge for many newcomers. This talk explains the underlying concepts of Bayesian modeling in an intuitive way without the math.

Intelligibility Throughout the Machine Learning Life Cycle

Jenn Wortman Vaughan, PhD

This session explores the importance of evaluating methods for achieving intelligibility in context with relevant stakeholders, ways of empirically testing whether intelligibility techniques achieve their goals, and why we should expand our concept of intelligibility beyond machine learning models to other aspects of machine learning systems, such as datasets and performance metrics.

Ai+

TRAINING

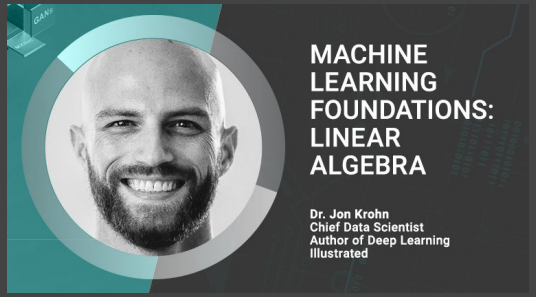
SELECTED SESSIONS

Machine learning is a sizable field that encompasses more than any single resource can cover. There are many tools, processes, languages, and platforms that will help you become a machine learning expert.

Below, we've highlighted just a few of the Machine Learning Ai+ Training sessions that will help you become a Machine Learning pro.

Ai+ TRAINING

SELECTED SESSIONS



Machine Learning Foundations: Linear Algebra

This first installment in the Machine Learning Foundations series the topic at the heart of most machine learning approaches. Through the combination of theory and interactive examples, you'll develop an understanding of how linear algebra is used to solve for unknown values in high-dimensional spaces, thereby enabling machines to recognize patterns and make predictions.

Supervised Machine Learning Series

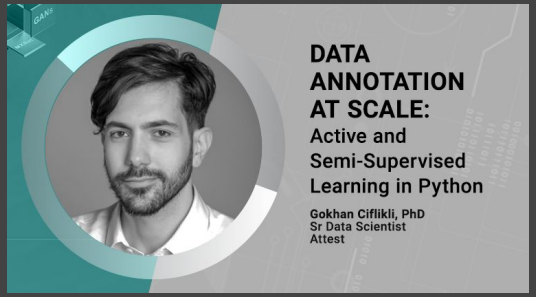
Walk through all steps of the classical supervised machine learning pipeline during this six-part series with Andras Zsom, PhD. During this course, you'll focus on topics like cross validation and splitting strategies, evaluation metrics, supervised machine learning algorithms, and interpretability.

Recommendation Systems in Python

Over the past few years, recommendation systems have become ubiquitous. We use them to select entertainment, financial products, and romantic partners. In this course, we will take a look at what you should consider when building your own recommendation system and how to get started building one using Python.

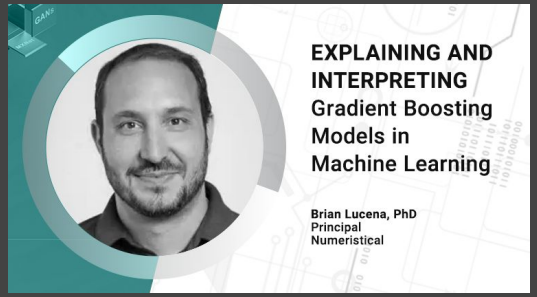
Ai+ TRAINING

SELECTED SESSIONS



Data Annotation at Scale: Active and Semi-Supervised Learning in Python

Explore how Active (human-in-the-loop) and Semi-Supervised (ML/AI-assisted) Learning frameworks can be combined to develop in-house solutions for executing rapid data labelling projects. By the end of the session, you will have a multitude of tools that you can utilize to scale up your data annotation efforts without losing all-important context.



Explaining and Interpreting Gradient Boosting Models in Machine Learning

Featuring hands-on practice using XGBoost with real-world data sets, this course will demonstrate how to approach data sets with the twin goals of prediction and understanding in a manner such that improvements in one area yield improvements in the other.

ODSC West 2020: Intelligibility Throughout the Machine Learning Lifecycle

This course will explore the importance of evaluating methods for achieving intelligibility in context with relevant stakeholders, ways of empirically testing whether intelligibility techniques achieve their goals, and why we should expand our concept of intelligibility beyond machine learning models to other aspects of machine learning systems, such as datasets and performance metrics.

ODSC UPCOMING EVENTS

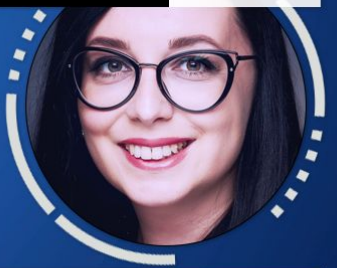
TIME SERIES FORECASTING WITH PYTHON BY MARTA MARKIEWICZ
Live Training
July 13th, 2021

Learn more

REINFORCEMENT LEARNING FOR GAME PLAYING AND MORE BY AMITA KAPOOR
Live Training
July 20th, 2021

Learn more

EXPLORING THE INTERCONNECTED WORLD: NETWORK/GRAPH ANALYSIS IN PYTHON BY NOEMI DERZSY
Live Training
August 3d, 2021

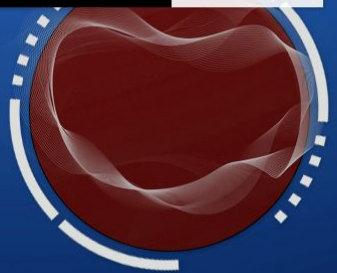


Learn more

BAYESIAN INFERENCE WITH PYMC BY ALLEN DOWNEY
Live Training
August 17th, 2021

Learn more

ODSC APAC 2021
Virtual Conference
September 15th – 16th, 2021



Learn more

ODSC WEST 2021
San Francisco
November 16th – 18th, 2021

Learn more

CONNECT WITH US

Ai+ Training

Open Data Science Blog

Data Science Job Board

ODSC Events

More Downloadable Guides:

Did you like this guide? We also have downloadable guides for deep learning and NLP. Download them for free now!

Webinars:

We offer free webinars several times a month, covering a variety of topics. Follow this page to learn more about upcoming webinars.

Weekly Newsletter:

Don't miss any future articles on data science and machine learning! Sign up for our weekly newsletter and get tutorials, insights, and the latest news sent to you directly.

Host Your Own Virtual Event

With the eventX.ai platform, you can host your own virtual events! Learn more here and schedule a demo.

Becoming a Part of ODSC Events:

Are you a technical or business expert in the world of data science and AI? Consider speaking at one of our events! Each event has its own speaker submission page:

ODSC APAC 2021 Virtual Conference
ODSC West 2021 Hybrid Conference

We also offer partnership opportunities!

Have your product, service, or research seen by thousands of data scientists at an event. **Learn more here.**