

Browse &gt; Data Science &gt; Machine Learning

This course is part of the Machine Learning Engineering for Production (MLOps) Specialization

## Machine Learning Modeling Pipelines in Production

★★★★★ 4.6 133 ratings • 24 reviews



Robert Crowe

[Go To Course](#)

Already enrolled

8,361 already enrolled

Offered By

 DeepLearning.AI

[About](#) [Instructors](#) [Syllabus](#) [Reviews](#) [Enrollment Options](#) [FAQ](#)

### About this Course

87,277 recent views

In the third course of Machine Learning Engineering for Production Specialization, you will build models for different serving environments; implement tools and techniques to effectively manage your modeling resources and best serve offline and online inference requests; and use analytics tools and performance metrics to address model fairness, explainability issues, and mitigate bottlenecks.

Understanding machine learning and deep learning concepts is essential, but if you're looking to build an effective AI career, you need production engineering capabilities as well. Machine learning engineering for production combines the foundational concepts of machine learning with the functional expertise of modern software development and engineering roles to help you develop production-ready skills.

Week 1: Neural Architecture Search  
 Week 2: Model Resource Management Techniques  
 Week 3: High-Performance Modeling  
 Week 4: Model Analysis  
 Week 5: Interpretability

#### WHAT YOU WILL LEARN

- Apply techniques to manage modeling resources and best serve batch and real-time inference requests.
- Use analytics to address model fairness, explainability issues, and mitigate bottlenecks.

#### SKILLS YOU WILL GAIN

[Explainable AI](#) [Fairness Indicators](#) [automl](#) [Model Performance Analysis](#) [Precomputing Predictions](#)

#### Flexible deadlines

Reset deadlines in accordance to your schedule.

#### Shareable Certificate

Earn a Certificate upon completion

#### 100% online

Start instantly and learn at your own schedule.

#### Course 3 of 4 in the

Machine Learning Engineering for Production (MLOps) Specialization

#### Advanced Level

- Some knowledge of AI / deep learning
- Intermediate Python skills
- Experience with any deep learning framework (PyTorch, Keras, or TensorFlow)

#### Approx. 26 hours to complete

#### English

Subtitles: English

### Instructor

Instructor rating 4.55/5 (40 Ratings) [\(1\)](#)

Robert Crowe

Instructor

TensorFlow Developer Engineer, Google

16,475 Learners

3 Courses

Offered by



DeepLearning.AI

DeepLearning.AI is an education technology company that develops a global community of AI talent.

DeepLearning.AI's expert-led educational experiences provide AI practitioners and non-technical professionals with the necessary tools to go all the way from foundational basics to advanced application, empowering them to build an AI-powered future.



When I left for maternity leave, I lost my interior design job. Health courses on Coursera helped me become a registered nurse and start a new path in life.

— Dimitra T.



Learning from leading scientists about what's going on in the field right now is so much different than the experience of reading a textbook.

— Peter W.



Coursera has been invaluable for my business personal development. The courses help my day-to-day problem solving.



### Other courses in this Specialization

**Introduction to Machine Learning in Production**



Introduction to Machine Learning in Production  
DeepLearning.AI

1 COURSE

**Machine Learning Data Lifecycle in Production**



Machine Learning Data Lifecycle in Production  
DeepLearning.AI

1 COURSE

**Deploying Machine Learning Models in Production**



Deploying Machine Learning Models in Production  
DeepLearning.AI

1 COURSE

### Syllabus - What you will learn from this course

WEEK

 7 hours to complete

1

#### Week 1: Neural Architecture Search

Learn how to effectively search for the best model that will scale for various serving needs while constraining model complexity and hardware requirements.



9 videos (Total 40 min), 2 readings, 6 quizzes [SEE LESS](#)



9 videos

Course Overview 2m

Hyperparameter Tuning 3m

Keras Autotuner Demo 6m

Intro to AutoML 5m

Understanding Search Spaces 2m

Search Strategies 4m

Measuring AutoML Efficacy 3m

AutoML on the Cloud 8m

Assignment Setup 1m



2 readings

Neural Architecture Search 3m

AutoML 3m



2 practice exercises

Hyperparameter Tuning and Neural Architecture Search 10m

AutoML 12m

WEEK



5 hours to complete

2

#### Week 2: Model Resource Management Techniques

Learn how to optimize and manage the compute, storage, and I/O resources your model needs in production environments during its entire lifecycle.

13 videos (Total 91 min), 3 readings, 3 quizzes [SEE LESS](#)

---

**13 videos**

- Dimensionality Effect on Performance 8m
  - Curse of Dimensionality 9m
  - Curse of Dimensionality: an Example 4m
  - Manual Dimensionality Reduction 5m
  - Manual Dimensionality Reduction: Case Study 7m
  - Algorithmic Dimensionality Reduction 2m
  - Principal Components Analysis 7m
  - Other Techniques 7m
  - Mobile, IoT, and Similar Use Cases 7m
  - Benefits and Process of Quantization 8m
  - Post Training Quantization 4m
  - Quantization Aware Training 4m
  - Pruning 12m
- 

**3 readings**

- Dimensionality Reduction Techniques 3m
  - Quantization 3m
  - Pruning 3m
- 

**2 practice exercises**

- Dimensionality Reduction 20m
  - Quantization and Pruning 20m
- 

WEEK

**5 hours to complete**

**3**

**Week 3: High-Performance Modeling**

Implement distributed processing and parallelism techniques to make the most of your computational resources for training your models efficiently.

6 videos (Total 57 min), 2 readings, 4 quizzes [SEE LESS](#)

---

**6 videos**

- Distributed Training 10m
  - High-Performance Ingestion 11m
  - Training Large Models - The Rise of Giant Neural Nets and Parallelism 13m
  - Teacher and Student Networks 3m
  - Knowledge Distillation Techniques 9m
  - Case Study - How to Distill Knowledge for a Q&A Task 8m
- 

**2 readings**

- High-Performance Modeling 5m
  - Knowledge Distillation 2m
- 

**2 practice exercises**

- High-Performance Modeling 20m
  - Knowledge Distillation 20m
- 

WEEK

**6 hours to complete**

**4**

**Week 4: Model Analysis**

Use model performance analysis to debug and remediate your model and measure robustness, fairness, and stability.



12 videos (Total 69 min), 5 readings, 5 quizzes

[SEE LESS](#)

#### ▶ 12 videos

- Model Performance Analysis 7m
- Introduction to TensorFlow Model Analysis 6m
- TFMA in Practice 3m
- Model Debugging Overview 3m
- Benchmark Models 1m
- Sensitivity Analysis and Adversarial Attacks 9m
- Adversarial Attack Demo 4m
- Residual Analysis 2m
- Model Remediation 4m
- Fairness 4m
- Measuring Fairness 5m
- Continuous Evaluation and Monitoring 14m

#### ⌚ 5 readings

- TensorBoard 3m
- TensorFlow Model Analysis 3m
- Sensitivity Analysis and Adversarial Attacks 3m
- Model Remediation and Fairness 5m
- Continuous Evaluation and Monitoring 5m

#### 📝 3 practice exercises

- Model Analysis 8m
- Model Analysis and Debugging 18m
- Continuous Evaluation and Monitoring 8m

WEEK

## 5

#### ⌚ 3 hours to complete

#### Week 5: Interpretability

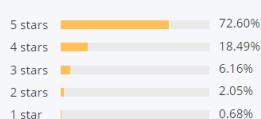
Learn about model interpretability - the key to explaining your model's inner workings to laypeople and expert audiences and how it promotes fairness and helps address regulatory and legal requirements for different use cases.



11 videos (Total 58 min), 9 readings, 4 quizzes

[SEE ALL](#)[Show Less](#)

## Reviews

**4.6** 24 reviews

#### TOP REVIEWS FROM MACHINE LEARNING MODELING PIPELINES IN PRODUCTION

by AK Aug 7, 2021

Some of the topics were too advanced and instructor assumes that we know those basics. It felt rush through little bit and more of reading slides than explaining at many places

by JS Sep 14, 2021

Excellent content and lectures from Mr. Robert . Thank you very much Sir for the excellent way of explaining these difficult topics . Thank you !!!

by MB Oct 21, 2021

I enjoyed this course a lot. It gave me a lot of ideas on how I can improve my models and make my workflow more efficient. Thank you.

by MT Sep 5, 2021

[View all reviews](#)

## About the Machine Learning Engineering for Production (MLOps) Specialization

Understanding machine learning and deep learning concepts is essential, but if you're looking to build an effective AI career, you need production engineering capabilities as well.

Effectively deploying machine learning models requires competencies more commonly found in technical fields such as software engineering and DevOps. Machine learning engineering for production combines

[SHOW ALL](#)



## Start Learning Today

- ✓ This Course Plus the Full Specialization
- ✓ Shareable Certificates
- ✓ Self-Paced Learning Option
- ✓ Course Videos & Readings
- ✓ Practice Quizzes
- ✓ Graded Assignments with Peer Feedback
- ✓ Graded Quizzes with Feedback
- ✓ Graded Programming Assignments

[Go To Course](#)

Shareable on LinkedIn



You can share your Course Certificates in the Certifications section of your LinkedIn profile, on printed resumes, CVs, or other documents.

**8,361** already enrolled

## Frequently Asked Questions

When will I have access to the lectures and assignments?

What will I get if I subscribe to this Specialization?

Is financial aid available?

More questions? Visit the [Learner Help Center](#).

### Coursera

- About
- What We Offer
- Leadership
- Careers
- Catalog
- Coursera Plus
- Professional Certificates
- MasterTrack® Certificates
- Degrees
- For Enterprise
- For Government
- For Campus
- Become a Partner
- Coronavirus Response

### Community

- Learners
- Partners
- Developers
- Beta Testers
- Translators
- Blog
- Tech Blog
- Teaching Center

### More

- Press
- Investors
- Terms
- Privacy
- Help
- Accessibility
- Contact
- Articles
- Directory
- Affiliates

### Mobile App

