

Machine Learning in Production

Dr. Jane Smith

Tech Conference 2024

Speaker Notes - Slide 1

Introduction points:

- Welcome audience and introduce topic
- Brief overview of ML challenges in production
- Agenda: models, monitoring, scaling, best practices
- Personal experience: 5 years ML engineering

Key statistics to mention:

- 85% of ML models never make it to production
- Average time from model to production: 8 months
- 60% of companies struggle with model monitoring

- Timing: 3 minutes
- Goal: Set context and engage audience
- Don't: Get too technical in introduction
- Do: Use relatable examples and ask questions

Challenges in ML Production

1

Model Drift

Performance degrades over time

2

Data Pipeline Failures

Upstream data quality issues

3

Scalability

Handling increasing load

Speaker Notes - Slide 2

Detailed talking points:

Model Drift:

- py" Real-world example: recommendation system accuracy
- py" Causes: changing user behavior, seasonal patterns
- py" Solutions: continuous monitoring, automatic retraining

Data Pipeline Failures:

- py" Story: 3am incident at previous company
- py" Impact: wrong predictions for 6 hours
- py" Prevention: data validation, circuit breakers

Scalability:

- py" Question for audience: "Who has seen 10x traffic spikes?"
- py" Technical solutions: caching, model optimization
- py" Architecture patterns: microservices, async processing