Survey : Challenges in Deploying Machine Learning: a Survey of Case Studies

1. INTRODUCTION

ML in business processes grows **25 percent year-over-year growth.**

Significant differences between **what works in an academic** setting and **what is required by a real-world system**.

8 and 90 days to deploy a single model, and 18% taking even more time

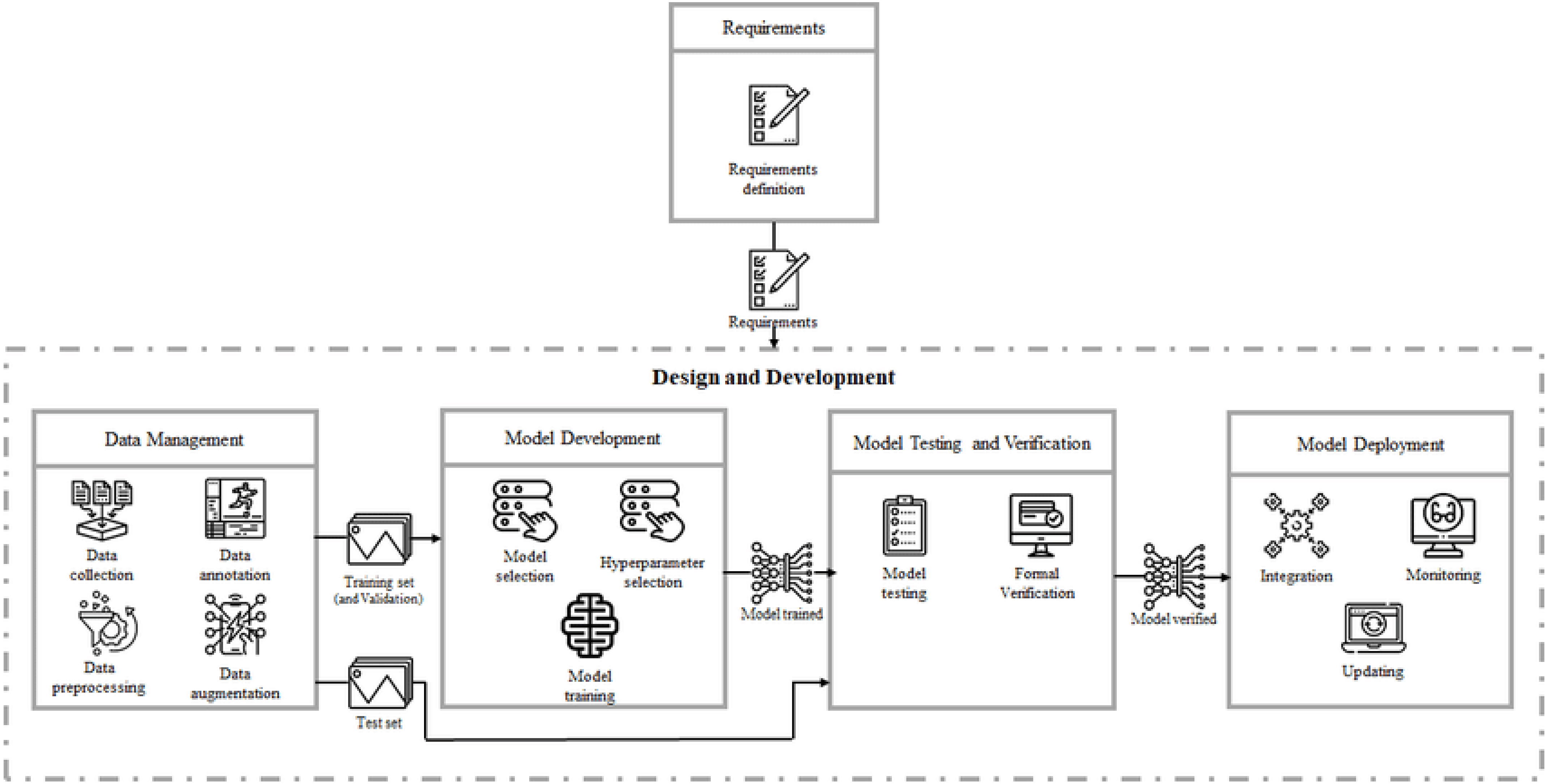
Significant portion of their attempted AI deployments fail, quoting

- **lack of expertise,**
- **bias in data,**
- **high costs**

In our survey, **three main types of papers** are considered.

- **Case studies,**
- **Rewiev papers,**
- **Lessons learned**

2. ML DEVELOPMENT WORKFLOW Ashmore et. al.



2. ML DEVELOPMENT WORKFLOW Ashmore et. al.

Deployment Stage	Deployment Step	Considerations, Issues and Concerns
Data Management	Data collection	• Data discovery
	Data preprocessing	• Data dispersion • Data cleaning
	Data augmentation	• Labeling of large volumes of data • Access to experts • Lack of high variance data
	Data analysis	• Data profiling
Model Learning	Model selection	• Model complexity • Resource-constrained environments • Interpretability of the model
	Training	• Computational cost • Environmental impact • Privacy-aware training
	Hyper-parameter selection	• Resource-heavy techniques • Unknown search space • Hardware-aware optimization
Model Verification	Requirement encoding	• Performance metrics • Business driven metrics
	Formal verification	• Regulatory frameworks
	Test-based verification	• Simulation-based testing • Data validation routines • Edge case testing

Deployment Stage	Deployment Step	Considerations, Issues and Concerns
Model Deployment	Integration	• Operational support • Reuse of code and models • Software engineering anti-patterns • Mixed team Dynamics
	Monitoring	• Feedback loops • Outlier detection • Custom design tooling
	Updating	• Concept drift • Continuous delivery
Cross-cutting aspects	Ethics	• Aggravation of biases • Fairness and accountability • Authorship • Decision making
	Law	• Country-level regulations • Abiding by existing legislation • Focus on technical solution only
	End user trust	• Involvement of end users • User experience • Explainability score
	security	• Data poisoning • Model stealing • Model inversion

3. DATA MANAGEMENT

Consequently, this stage consumes time and energy that is often not anticipated beforehand.

Deployment Stage	Deployment Step	Considerations, Issues and Concerns
Data Management	Data collection	• Data discovery
	Data preprocessing	• Data dispersion • Data cleaning
	Data augmentation	• Labeling of large volumes of data • Access to experts • Lack of high variance data
	Data analysis	• Data profiling

3. DATA MANAGEMENT

3.1. Data Collection

Finding data sources and **understanding their structure** is a major task, which may prevent data scientists from even getting started on the actual application development.

Photo : eventsget.com

Photo : questionpro.com

3. DATA MANAGEMENT

3.2. Data Preprocessing

- ➢ Data dispersion

Multiple data sources with:

- Different **schemas**
- Different **conventions**
- Different ways of **storing and accessing data**

- ➢ Data cleaning

Multiple data sources with:

- Identification of a **schema**
- **Imputation**
- **Reduction** of data

3. DATA MANAGEMENT

3.3. Data Augmentation

In Survey:

Real-world data is often **unlabeled** because of:

- **limited access to experts**,
  - Medical Image analysis
- Absence of **high variance** data,
  - Especially for Reinforcement Learning (RL)
- **Sheer volume**
  - **Network Domain.** Two ways of collection data:
    - **Uncontrolled**, collecting **real** traffic,
    - **Controlled**, **emulating** or generatic traffic.

In Google:

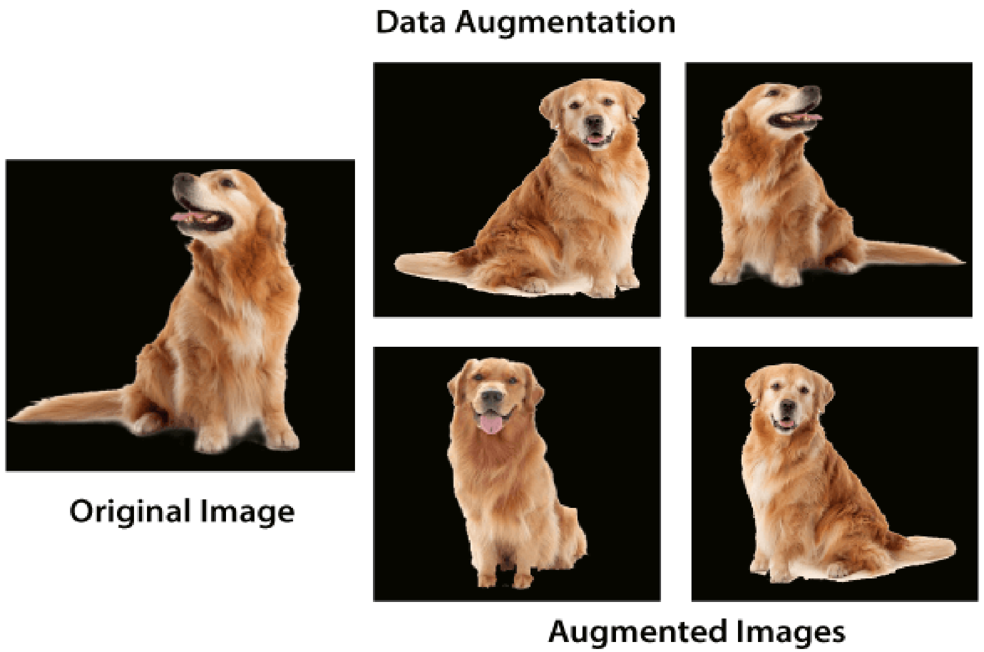


Photo: javatpoint.com

3. DATA MANAGEMENT

3.4. Data Analysis

Biggest challange: **Data visualisation**

- There are still **too few tools** for this.
- data issues are the **main reason to doubt the quality of the overall work.**

Photo : boostlabs.com

4. MODEL LEARNING

Deployment Stage	Deployment Step	Considerations, Issues and Concerns
Model Learning	Model selection	• Model complexity • Resource-constrained environments • Interpretability of the model
	Training	• Computational cost • Environmental impact • Privacy-aware training
	Hyper-parameter selection	• Resource-heavy techniques • Unknown search space • Hardware-aware optimization

4. MODEL LEARNING

4.1 Model Selection

Maind tradeoff: **High accuracy** vs **low complexity(simpler models)**

In practice, **simpler models are often** choosen such as **Shallow NN, PCA, DT, Random Forests.**

Simpler models ease the followings:

- **proving concept,**
- **end-to-end solution time,**
- **moderate hardware**
- **Interpretability**

Some domains using simpler models

- Wireless cellular networks: Limited **energy, memory, and data transmission**
- Banking Industry: **Interpretability**
- UAV: Requires complex models but computatiaonal resource demands still blocks **online processing**.

4. MODEL LEARNING

4.2 Model Training

- **Costs:**
  - **Economic costs**
  - **Computational resources** required

  Often true in **NLP**

- **Environmental Impacts:**

ML model training is driving up **energy consumption** and greenhouse **gas emissions.**

- **Privacy Aware Training:**

Two concerns:

  - **Privacy of data**
  - **Preservation of sensitive data.**

Tradeoff between **privacy** and **utility.**

- There are two methods to overcome this tradeoff:
  - **Homomorphic encyrpition**
  - **Federated learning**

4. MODEL LEARNING

4.3 Hyper-Parameter Selection

Computationally challenging because size of the HPO task grows **exponentially**.

- **Hardware-aware ML**: One needs to be aware of **energy** and **memory constraints** imposed by mobile and embedded devices.

5. MODEL VERIFICATION

Deployment Stage	Deployment Step	Considerations, Issues and Concerns
Model Verification	Requirement encoding	• Performance metrics • Business driven metrics
	Formal verification	• Regulatory frameworks
	Test-based verification	• Simulation-based testing • Data validation routines • Edge case testing

5. MODEL VERIFICATION

5.1. Requirement Encoding

Increase in **model performance** does not translate into a gain in **business value**.

- Additional **domain specific metrics** need to be defined and measured.
  - A **cross-disciplinary effort** is needed

5.2. Formal Verification

Verifying that **software functionality** follows the **requirements** defined within **the scope of the project.**

5. MODEL VERIFICATION

5.3. Test Based Verification

Ensuring that the model generalizes well to previously **unseen data.**

Full scale testing in a real-world environment can be challenging for a variety of

- **safety**,
- **security** and
- **scale** reasons,
  - ➢ It is often substituted with testing in **simulation**. Use of simulations is a de-facto standard in RL for training agents

In addition:

Dataset itself also needs to be **constantly validated**

Data issues can originate from

- **bugs in code,**
- **feedback loops,**
- **changes in data** dependencies**.**

6. MODEL DEPLOYMENT

Deployment Stage	Deployment Step	Considerations, Issues and Concerns
Model Deployment	Integration	• Operational support • Reuse of code and models • Software engineering anti-patterns • Mixed team Dynamics
	Monitoring	• Feedback loops • Outlier detection • Custom design tooling
	Updating	• Concept drift • Continuous delivery

6. MODEL DEPLOYMENT

Production systems are **complex software systems** that have to be **maintained** over time.

DevOps: Discipline that focuses on **techniques** and **tools** required to successfully **maintain** and **support** existing **production systems.**

There is a neccessity to apply DevOps principles to ML systems.

- **AIOps**

6. MODEL DEPLOYMENT

6.1. Integration

- **Operational Support**

Two main activities

- **Building the infrastructure** to run the model and
- **Implementing the model itself** in a form that can be **consumed** and **supported**.

- **Reuse of codes and models**

Use of the **same codes or models** can be challenging **in different systems.**

- **Software engineering anti-patterns**

ML is used in cases where the software has to take an explicit **dependency on external data**

➢ Anti-patterns are widely used in ML softwares

6. MODEL DEPLOYMENT

6.1. Integration

- **Mixed team dynamics**

Researchers and software **engineers** often **works together** on the same project.
There is **NO clear separation of responsibilities.**

- Contributors in **both roles** often work on the **same code.**

6. MODEL DEPLOYMENT

6.2. Monitoring

Monitoring is required for **maintaining** of ML systems.

- What are the **key metrics of data and models** to monitor and how to trigger system alarms?

Monitoring of folowings is an open problem

- **evolving input data**,
- **prediction bias** and
- **overall performance** of ML models

6. MODEL DEPLOYMENT

6.2. Monitoring

Three main issues in model maintanenece:

- **Feed-back loops**

ML models in production can **influence their own behavior** over time.

  - ➢ Tradeoff between **staying up to date** vs **Feedback loops**

- **Outlier detection**

Labeled outlier data is scarce.

  - ➢ Problem turns into a **semisupervised** or even **unsupervised** problem.

- **Custom tooling**

Out-of-the-box tooling does **not fit** projects **spesific needs** well.

6. MODEL DEPLOYMENT

6.3. Updating

Two techniques for **adapting models to new data:**

- **Regular training**
- **Continual learning**

Model updating is affected by practical considerations

- **Concept drift** (aka dataset drift)

Changes observed in **joint distribution**.

- Can be caused by an **inability to avoid fluctuations** in the data collection procedure
- **Microscopic shifts → Noticebla consequences**

6. MODEL DEPLOYMENT

6.3. Updating

- **Continuous delivery (CD)**
  - How to deliver the model artifact to the production environment?

CD for machine learning solutions is **complicated** because unlike regular software projects, ML solutions experience change along three axes:

- **the code,**
- **the model,**
- **the data.**

7. CROSS-CUTTING ASPECTS

Deployment Stage	Deployment Step	• Considerations, Issues and Concerns
Cross-cutting aspects	Ethics	• Aggravation of biases • Fairness and accountability • Authorship • Decision making
	Law	• Country-level regulations • Abiding by existing legislation • Focus on technical solution only
	End user trust	• Involvement of end users • User experience • Explainability score
	security	• Data poisoning • Model stealing • Model inversion

7. CROSS-CUTTING ASPECTS

7.1 Ethics

ML models can rely on **hidden biases** that already exist in data.

Examples:

- Facial analysis: In a facial analysis model, darker skinned females are the most misclassified group since dataset is imbalanced on the basis of skin color.
- Creative Arts: When a trained model is used to create a piece of visual art, it is not entirely clear where the authorship of this piece resides.

7. CROSS-CUTTING ASPECTS

7.2 Laws

Various countries have produced **regulations** to **protect personal data** rights.

- More sensitive data → Stronger regulations
  - ➢ Adoption of ML in **healthcare** is particularly difficult.

Legislation takes time to develop.

- ➢ Cannot keep up with the speed of ML.

7. CROSS-CUTTING ASPECTS

7.3 End Users' Trust

ML is often met cautiously by the end users.

There are three considerations to get end users' trust:

- **Involvement of end users**

Getting the users involved in the early steps of the project helps feel them confident, especially in **medicine**.

- **Explainability score**

Model **interpretability has limits** as a trust-building tool.

  - ➢ Other ways should be considered.

- **User experience**

Bad **user interface** → Obstacles in adoption of new technology

7. CROSS-CUTTING ASPECTS

7.4 Security

ML opens up opportunities for new types of **security attacks.** Attacks can ocur on

- model itself,
- training data,
- predicitons.

- **Data poisoning**

Corrupting the integrity of the model during the **training phase**

- Particularly relevant with systems which ML models **continuously updated** with **newly incoming data**.

- **Model stealing**

Querying model **inputs** and monitoring **outputs**.

- **Model inversion**

Recovering parts of the training set.

8. DISCUSSION OF POTENTIAL SOLUTIONS

Further **growth of ML adoption** can be severely hindered by **poor deployment** experience.

It is critical to **understand critical pain points** and provide

- **tools,**
- **services,**
- **best practices.**

Possible research avenues are categorized into two

- **Tools and Services**
- **Holistic Approach**

8. DISCUSSION OF POTENTIAL SOLUTIONS

8.1. Tools and services

Tools can be improved to develop following utilities.

- **Data storage facility**
- **Model hosting with APIs** for training and inference operations
- **Common metrics** to monitor model health
- **Interface** to accept custom changes from the user
- **Quality assurance**
- **Checklist methodology**
- **Weak supervision** (Snorkel, Snuba, cleanlab)
- **AutoML** (Auto-Keras, Auto-sklearn, TPOT)
- **Detection of mitigation of unnoticed dataset shift** (Alibi Detect, services Azure ML, AWS Sagemaker)

8. DISCUSSION OF POTENTIAL SOLUTIONS

8.1. Tools and services

Using a **particular tool** in the project → Additional **dependency** to that tool

The **more tools** used in the project → The **more dependency**

- **Management** becomes a problem.

8. DISCUSSION OF POTENTIAL SOLUTIONS

8.2. Holistic approaches

Managing an ML project is not similar to managing a regular software project.
They do not fit well to common management processes like Scrum or Waterfall.

The main differences arise from unique activities like

- data discovery,
- dataset preparation,
- model training,
- deployment success measurement, etc.

8. DISCUSSION OF POTENTIAL SOLUTIONS

8.2. Holistic approaches

Some considerations:

- **datasheets for datasets**

Makes data collection and management easier.

- **Data Oriented Architecture (DOA)**

Makes data collection and management easier.

- **set of guidelines** and **best practices**

Helps developers make right decisions.

  - The Association of German Engineers (VDI) has released a **series of guidelines on various aspects of big data applications** in the manufacturing industry.
  - Zinkevich compiled a **collection of best practices** for machine learning that are utilized in Google.

It should be noted that

All such approaches assume **significant time investment**, because they represent **significant changes to current norms in project management and development.**