



Title: Data Governance

Course: Machine Learning

Instructor: Claudio Sartori

Master: Data Science and Business Analytics

Master: Artificial Intelligence and Innovation Management

Academic Year: 2024/2025

# Introduction

- Data governance ensures data is managed properly.
- It involves policies, procedures, and controls.
- Objective: Maintain data quality, security, and compliance.

# Key Components of Data Governance

- **Data Quality:** Ensuring accuracy, consistency, and reliability.
- **Data Security:** Protecting data from unauthorized access.
- **Data Privacy:** Compliance with regulations (e.g., GDPR, HIPAA).
- **Data Stewardship:** Assigning roles for data responsibility.
- **Metadata Management:** Organizing data definitions and lineage.

# Benefits of Data Governance

- Improved decision-making.
- Enhanced regulatory compliance.
- Better data security and risk management.
- Increased operational efficiency.
- Higher trust in data quality.

# Challenge: Lack of Executive Support

- Difficulty in securing funding and organizational buy-in.
- Leadership may not see immediate ROI from data governance efforts.
- Requires alignment with business objectives to gain support.

# Challenge: Resistance to Change

- Employees may be reluctant to adopt new policies and procedures.
- Change management strategies are needed for smooth adoption.
- Training and awareness programs help mitigate resistance.

# Challenge: Complexity of Data Ecosystems

- Organizations handle vast amounts of diverse data.
- Data is often stored across multiple platforms and formats.
- Integration and consistency are key challenges.

# Challenge: Ensuring Continuous Compliance

- Regulatory requirements (e.g., GDPR, HIPAA) evolve over time.
- Organizations must continuously monitor and update policies.
- Automation tools can assist in maintaining compliance.



# Data Quality

- Ensures that data is accurate, complete, and reliable.
- Affects decision-making, operational efficiency, and compliance.
- Poor data quality can lead to misinformation and inefficiencies.

# Measuring Data Quality

- **Accuracy:** Data should reflect real-world values correctly.
- **Completeness:** No missing or incomplete records.
- **Consistency:** Data should be uniform across different sources.
- **Timeliness:** Data should be up to date and relevant.
- **Validity:** Data must conform to defined formats and rules.
- **Uniqueness:** No duplicate records within the dataset.

# Data Quality Assessment Checklist

- **Accuracy:** Does the data correctly represent real-world entities?
- **Completeness:** Are all required fields filled? Any missing values?
- **Consistency:** Is the data uniform across different sources and formats?
- **Timeliness:** Is the data up-to-date and relevant for its intended use?
- **Validity:** Does the data conform to the required format and constraints?
- **Uniqueness:** Are there duplicate records that need to be removed?
- **Integrity:** Are relationships between data entities correctly maintained?

# Conclusion

- Data governance is crucial for modern organizations.
- It ensures data is secure, reliable, and compliant.
- Strong governance frameworks enhance business value.