Data Fundamentals for AI/ML

Surendra Panpaliya

Agenda

- Importance of Data in AI/ML
- Data Collection and Preprocessing Power/Energy Sector
- Handling missing data
- Feature engineering and selection.
- Overview of Structured vs. Unstructured Data

Importance of Data in AI/ML

Power/Energy Sector

Accurate Decision-Making for Energy Management

Role of Data:

Data from energy production,

consumption, and distribution

enables predictive and

prescriptive analytics.

Example

REAL-TIME MONITORING OF POWER GRIDS

HELPS OPTIMIZE ENERGY DISTRIBUTION AND

PREVENT BLACKOUTS.

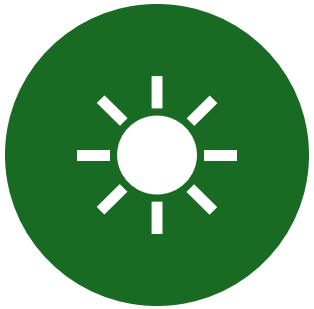
Impact

REDUCED ENERGY WASTE.

IMPROVED GRID STABILITY AND EFFICIENCY.

Renewable Energy Optimization

ROLE OF DATA



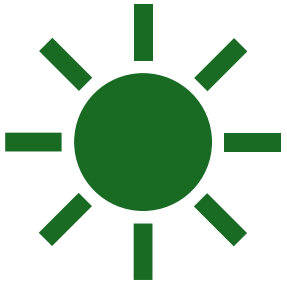
WEATHER, SOLAR RADIATION, WIND SPEED DATA

USED TO FORECAST

RENEWABLE ENERGY GENERATION.

Example

AI models

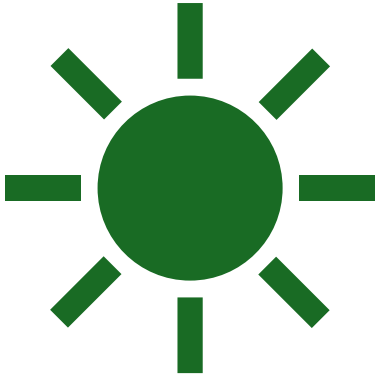


Predict solar panel output

Optimize energy storage systems.

Impact

Maximized use of renewable energy.



Reduced dependency on non-renewable sources.

Predictive Maintenance

Role of Data

Sensor data from

Machinery and Equipment

Used to predict failures.

Example

Analyzing vibration

Temperature data

To forecast

Turbine malfunctions.

Impact

Minimizes downtime.

Reduces maintenance costs.

Load Forecasting

Role of Data

Historical

Real-time consumption data

Predict future energy demands.

Example

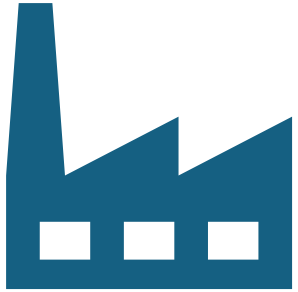
AI models use

Seasonal trends

External factors

like weather.

Impact



Better alignment of energy supply and demand.

Cost savings in energy production.

Environmental Compliance and Carbon Footprint Reduction

Role of Data

Data on emissions

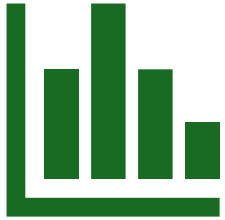
Energy usage

Helps monitor

Reduce environmental impact.

Example

AI models



Analyze emissions data

to ensure regulatory compliance.

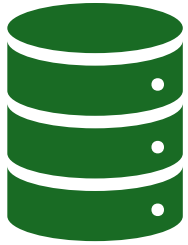
Impact

Improved sustainability practices.

Enhanced corporate social responsibility.

Challenges in Using Data

Data Quality



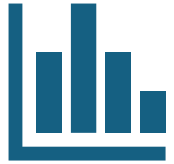
Data Integration

Data Security

Data Quality

Incomplete,

inconsistent, or

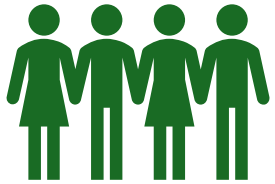


biased data

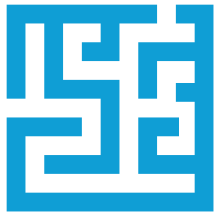
can affect AI/ML outcomes.

Data Integration

Combining data



from diverse sources



is complex.

Data Security

PROTECTING

SENSITIVE DATA

IN COMPLIANCE

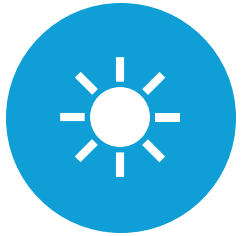
WITH REGULATIONS

IS CRITICAL.

Data in Power/Energy Sector

DATA DRIVES EFFICIENT ENERGY USE

SUPPORTS GRID LOAD FORECASTING



OPTIMIZES RENEWABLE ENERGY RESOURCES

IMPROVES PREDICTIVE MAINTENANCE SYSTEMS

ENHANCES DECISION-MAKING PROCESSES

Smart Grid Applications

Data enables real-time monitoring

Improves grid stability and reliability

Supports outage detection and response

Facilitates energy demand management

Drives smarter infrastructure solutions

Renewable Energy Optimization

- AI uses data for forecasts
- Improves solar and wind efficiency
- Balances grid supply-demand mismatch
- Reduces energy wastage significantly
- Promotes sustainable energy practices

Predictive Maintenance in Energy

MONITORS EQUIPMENT HEALTH PROACTIVELY

DATA PREDICTS FAILURES BEFORE OCCURRENCE

REDUCES DOWNTIME AND MAINTENANCE COSTS

ENHANCES ASSET LIFE AND RELIABILITY

IMPROVES OPERATIONAL EFFICIENCY

Data in IT Operations

Data enhances IT system insights

Improves service reliability and uptime

Supports proactive issue resolutions

Drives efficient resource utilization

Enables intelligent automation processes

Network Performance Optimization

- Analyzes traffic patterns with data
- Prevents bottlenecks and service issues
- Ensures optimal bandwidth allocation
- Supports adaptive network configurations
- Improves user experience significantly

Cybersecurity Enhancement

Detects threats using data patterns

Enhances real-time incident response

Predicts vulnerabilities and mitigations

Supports compliance with regulations

Builds robust defense mechanisms

AI-Driven IT Automation

Automates repetitive IT tasks

Reduces manual intervention efforts

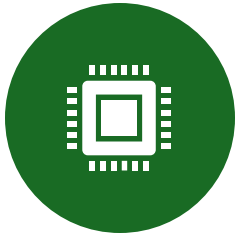
Optimizes resource allocation dynamically

Improves operational cost-efficiency

Accelerates business transformation

Importance of Data Collection

DATA DRIVES INSIGHTS AND DECISIONS



COLLECTS REAL-TIME SYSTEM PERFORMANCE

ENABLES ENERGY DEMAND FORECASTING

SUPPORTS GRID LOAD MANAGEMENT

IMPROVES ENERGY EFFICIENCY STRATEGIES

Sources of Data Collection

SMART METERS AND IOT SENSORS

WEATHER AND ENVIRONMENTAL DATA

GRID PERFORMANCE MONITORING SYSTEMS

ENERGY CONSUMPTION TRACKING DEVICES

RENEWABLE ENERGY GENERATION DATA

Challenges in Data Collection

HANDLING DIVERSE DATA SOURCES

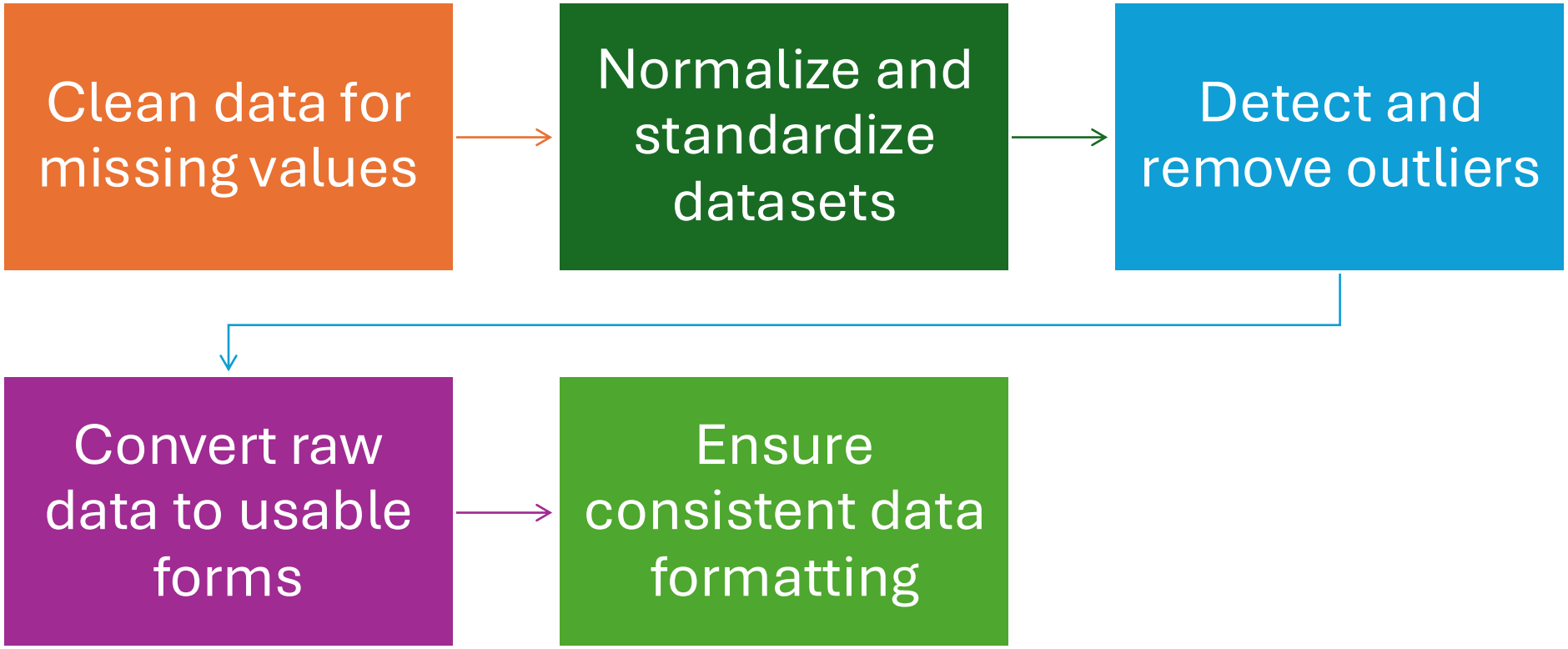
ENSURING REAL-TIME DATA ACCURACY

INTEGRATING LEGACY AND NEW SYSTEMS

MANAGING MASSIVE DATA VOLUMES

ENSURING DATA SECURITY COMPLIANCE

Data Preprocessing Steps



Impact of Preprocessing

Improves model training efficiency

Enhances data-driven decision-making

Reduces computational resource usage

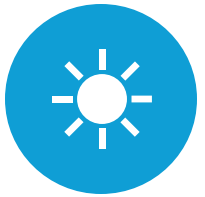
Increases reliability of predictions

Supports actionable insights extraction

Importance of Data in Energy

Data optimizes energy utilization

Enables predictive grid management



Supports renewable energy forecasting

Enhances energy efficiency strategies

Reduces operational costs significantly

Applications in Smart Grids

Real-time grid performance tracking

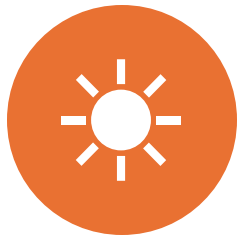
Prevents outages with quick response

Optimizes load distribution effectively

Supports energy demand predictions

Improves grid stability and reliability

Impact on Renewable Energy



FORECASTS SOLAR AND WIND OUTPUTS

BALANCES ENERGY SUPPLY AND DEMAND

MINIMIZES RENEWABLE ENERGY WASTAGE

SUPPORTS SUSTAINABLE ENERGY INTEGRATION

BOOSTS OVERALL ENERGY RELIABILITY

Role of Data in IT Ops

ENHANCES SYSTEM PERFORMANCE MONITORING

IMPROVES INCIDENT RESOLUTION ACCURACY

SUPPORTS PROACTIVE ISSUE DETECTION

OPTIMIZES RESOURCE ALLOCATION DYNAMICALLY

ENABLES INTELLIGENT AUTOMATION PROCESSES

Data-Driven Cybersecurity

ANALYZES PATTERNS FOR THREAT DETECTION

IMPROVES RESPONSE TO CYBER INCIDENTS

PREVENTS BREACHES WITH PREDICTIVE MODELS

SUPPORTS ROBUST SECURITY FRAMEWORKS

BUILDS TRUST THROUGH SECURE OPERATIONS

Optimizing IT Infrastructure

MONITORS AND PREDICTS USAGE TRENDS

REDUCES DOWNTIME WITH SMART INSIGHTS

SUPPORTS SCALING WITH ACCURATE FORECASTS

ENHANCES CLOUD RESOURCE MANAGEMENT

IMPROVES COST EFFICIENCY SIGNIFICANTLY

Role of Data in IT Ops

Data ensures operational efficiency

Supports predictive system maintenance

Improves resource utilization dynamically

Enhances decision-making with insights

Drives innovation in IT systems

Proactive Issue Resolution

AI DETECTS ANOMALIES EARLY

PREDICTS SYSTEM FAILURES PROACTIVELY

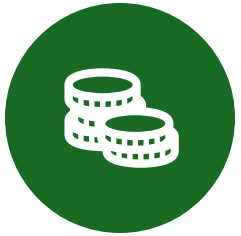
REDUCES DOWNTIME SIGNIFICANTLY

ENSURES SEAMLESS IT OPERATIONS

ENHANCES USER SATISFACTION LEVELS

Optimizing Resource Allocation

MONITORS RESOURCE USAGE PATTERNS



IMPROVES COST MANAGEMENT STRATEGIES

SUPPORTS DYNAMIC SCALING SOLUTIONS

OPTIMIZES HARDWARE AND SOFTWARE RESOURCES

DRIVES SUSTAINABLE IT PRACTICES

Enhancing IT Security

ANALYZES THREATS USING DATA MODELS

PREVENTS BREACHES WITH REAL-TIME ALERTS

ENSURES COMPLIANCE WITH REGULATIONS

SUPPORTS ROBUST SECURITY MEASURES

BUILDS TRUST IN IT INFRASTRUCTURE

Data-Driven Automation

AUTOMATES REPETITIVE IT TASKS

IMPROVES EFFICIENCY AND ACCURACY

REDUCES MANUAL INTERVENTION EFFORTS

ENABLES SELF-HEALING IT SYSTEMS

SUPPORTS CONTINUOUS SYSTEM MONITORING

Data Quality and Availability

AI needs high-quality data

Incomplete data leads to errors

Challenges in accessing real-time data

Ensuring data accuracy is critical

Overcome missing or biased data

Computational Power Requirements

AI DEMANDS SIGNIFICANT PROCESSING POWER

HIGH COSTS FOR ADVANCED HARDWARE

LIMITS ACCESSIBILITY FOR SMALL FIRMS

OPTIMIZE ALGORITHMS FOR EFFICIENCY

ADOPT CLOUD SOLUTIONS STRATEGICALLY

Bias and Fairness

- AI models can inherit biases
- Skewed data causes unfair results
- Affects marginalized communities deeply
- Requires diverse training datasets
- Ensures fairness in decision-making

Explainability and Transparency

- AI models are often 'black boxes'
- Difficult to explain model outputs
- Transparency builds trust in systems
- Develop interpretable AI frameworks
- Stakeholders demand clear explanations

Ethical and Regulatory Issues

- Lack of universal AI ethics standards
- Challenges in privacy and compliance
- Prevent misuse of AI technologies
- Governments regulate AI development
- Balance innovation with accountability

Integration with Existing Systems

LEGACY SYSTEMS POSE INTEGRATION HURDLES

ENSURE SEAMLESS DATA FLOW PROCESSES

HIGH COSTS FOR SYSTEM UPGRADES

REQUIRES SKILLED PROFESSIONALS ONBOARD

MANAGE DISRUPTION DURING TRANSITIONS

Cybersecurity Threats

- AI systems vulnerable to attacks
- Adversarial attacks compromise reliability
- Securing models is a challenge
- Requires robust cybersecurity measures
- Protect sensitive information effectively

Talent and Skill Gaps

SHORTAGE OF SKILLED AI PROFESSIONALS

REQUIRES EXPERTISE IN MULTIPLE DOMAINS

UPSKILL WORKFORCE TO MEET DEMANDS

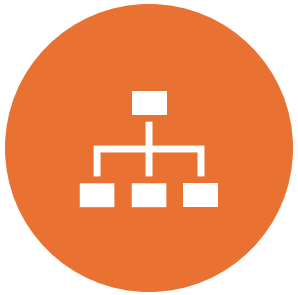
PROMOTE AI/ML EDUCATION WIDELY

ENCOURAGE INTERDISCIPLINARY COLLABORATIONS

Structured vs. Unstructured Data

Surendra Panpaliya

What is Structured Data?



Organized

Stored in a predefined format,

Often in rows and columns,

Making it easy to search, process, and analyze.

Structured Data Characteristics

Stored in databases like

SQL, Excel, or data warehouses.

Schema-based

Clearly defined data fields

Structured Data Characteristics

Easy to analyze

using traditional tools

like SQL or BI platforms.

Numerical or Categorical.

Examples in Power/Energy Sector

SENSOR READINGS

DATA FROM IOT DEVICES

MONITORING TEMPERATURE,

PRESSURE, ENERGY FLOW

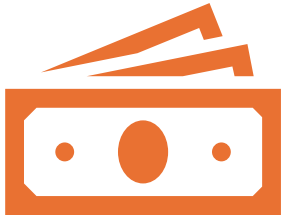
Examples in Power/Energy Sector

HISTORICAL DATA

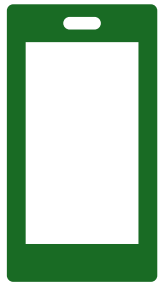
MONTHLY ENERGY PRODUCTION OR

CONSUMPTION RECORDS.

Examples in Power/Energy Sector



Billing Data



Customer usage

Payment records.

Use Cases

Load Forecasting

Predict energy demand

using structured time-series data.

Use Cases

Predictive Maintenance

Analyze equipment health

based on sensor readings.

What is Unstructured Data?

LACKS A PREDEFINED FORMAT,

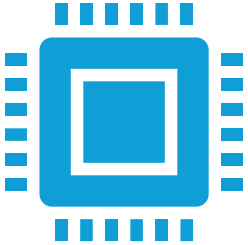
MAKING IT MORE CHALLENGING

TO STORE, SEARCH, AND ANALYZE.

What is Unstructured Data?

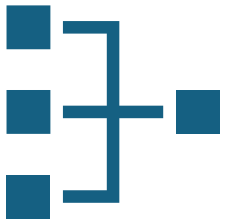
Requires advanced tools

like Natural Language Processing (NLP) or



Computer Vision.

Characteristics



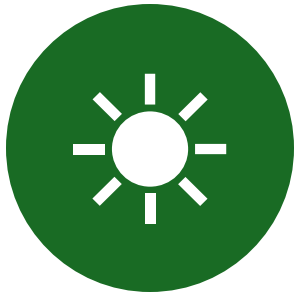
Does not follow a fixed schema.

Requires preprocessing to derive insights.

Can include text, images, videos, or audio.

Examples in Power/Energy Sector

Satellite Images



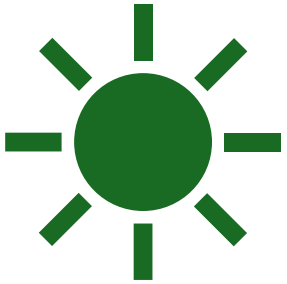
Monitor solar farms,

wind farms, or

grid infrastructure.

Examples in Power/Energy Sector

Identifying



Damaged Solar panels

from Aerial imagery

Examples in Power/Energy Sector

MAINTENANCE LOGS

FREE-TEXT REPORTS

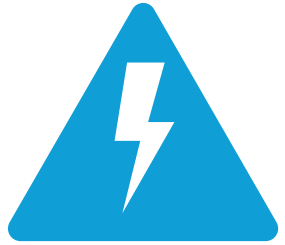
DESCRIBING

ISSUES OR REPAIRS.

Examples in Power/Energy Sector

Social Media Data

Customer complaints

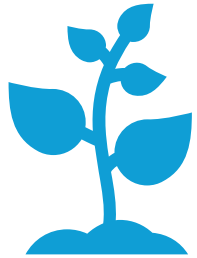


Feedback about power outages.

Examples in Power/Energy Sector

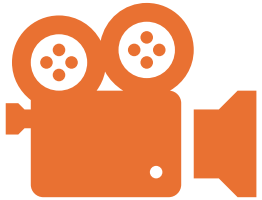
Audio Data

Technician calls or
Voice reports



from the field.

Examples in Power/Energy Sector



Video Footage

Surveillance of power plants

for security purposes.

Use Cases

Asset Monitoring

Use computer vision

to detect faults in power lines.

Use Cases

Customer Sentiment Analysis

Use NLP to analyze social media feedback

about outages or service quality.

Surendra Panpaliya

Founder and CEO

GKTCS Innovations

https://www.gktcs.com