Hochschule Emden/Leer
Abteilung Maschinenbau
MSR-Labor

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Walmart Sales Forecasting System

Web-Based Time Series Forecasting with Interactive Analytics

Hemanth Jadiswami Prabhakaran, Ayush Plawat, Adil Ibraheem Koyava

June 24, 2025

Problem Description & Motivation

Retail Sales Forecasting - The Challenge I

Setting the Context

The Scale of Modern Retail

- Walmart operates over **11,500 stores** across multiple countries
- Each store contains **dozens of departments** with unique sales patterns
- Weekly sales data creates over **4,400 individual time series**
- Annual revenue exceeds **$500 billion** globally

Why Accurate Forecasting Matters

- **Inventory Costs**: Poor forecasting leads to $1.1 trillion in excess inventory globally
- **Stockouts**: Lost sales opportunities cost retailers 4% of annual revenue

Retail Sales Forecasting - The Challenge II

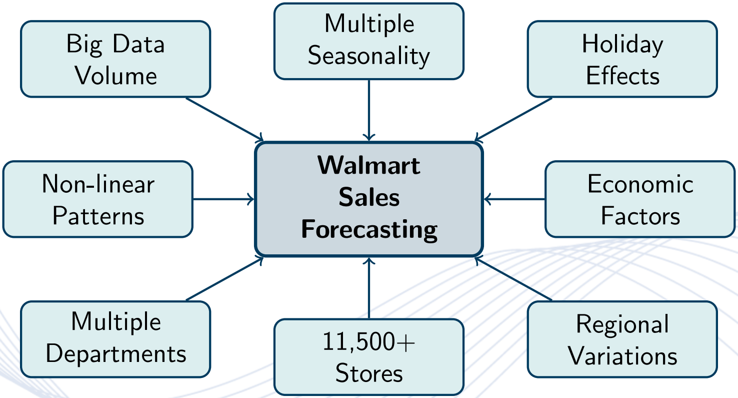
Setting the Context

- **Operational Efficiency**: Staff scheduling and resource allocation depend on demand predictions
- **Competitive Advantage**: Better forecasts enable superior customer service

Challenge: How do we predict complex, multi-seasonal retail patterns accurately?

The Forecasting Challenge I

Retail Sales Complexity



Walmart - World's Largest Retailer

The Forecasting Challenge II

Retail Sales Complexity

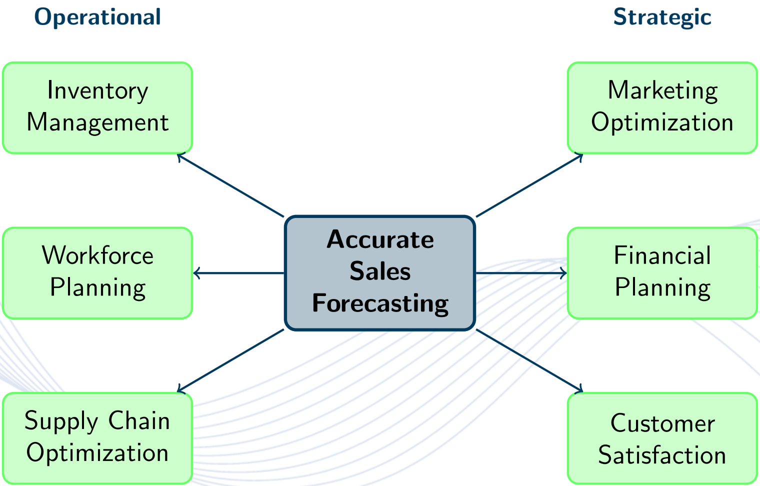
- 11,500+ stores worldwide [Zha21]
- Multiple departments per store
- Complex sales patterns with multiple seasonalities
- Influenced by holidays, economic factors, regional variations

Key Challenges

- Weekly and annual seasonal patterns
- Holiday effects (Christmas, Black Friday, Easter)
- Economic indicators impact
- Store-to-store variations

Business Impact I

Why Accurate Forecasting Matters



Business Impact II

Why Accurate Forecasting Matters

Operational Benefits

- **Inventory Management**: Optimal stock levels, reduced waste
- **Workforce Planning**: Efficient staff allocation
- **Supply Chain**: Better coordination with suppliers

Strategic Benefits

- **Marketing Optimization**: Campaign effectiveness evaluation
- **Financial Planning**: Revenue forecasting and budgeting
- **Customer Satisfaction**: Avoiding stockouts and overstocking

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Dataset Overview I

Walmart Sales Data Characteristics

Data Scope

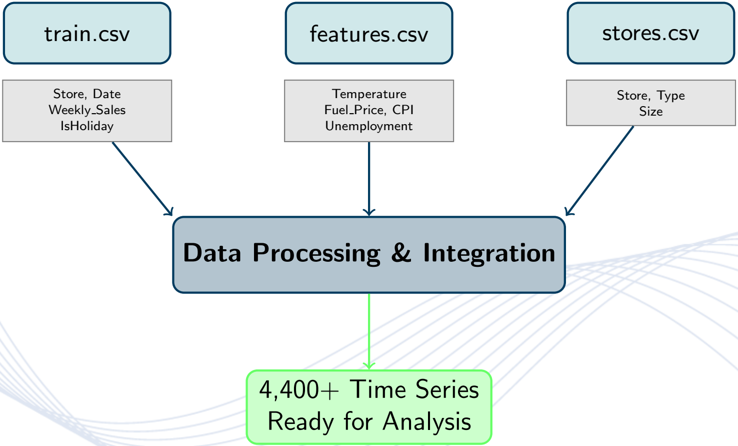
- **45 stores** across different regions
- **Multiple departments** per store
- **Weekly sales** from 2010-2012
- Over **4,400 time series** to analyze

External Factors

- Temperature, fuel prices
- Unemployment rates
- Consumer Price Index (CPI)
- Holiday indicators

Dataset Overview II

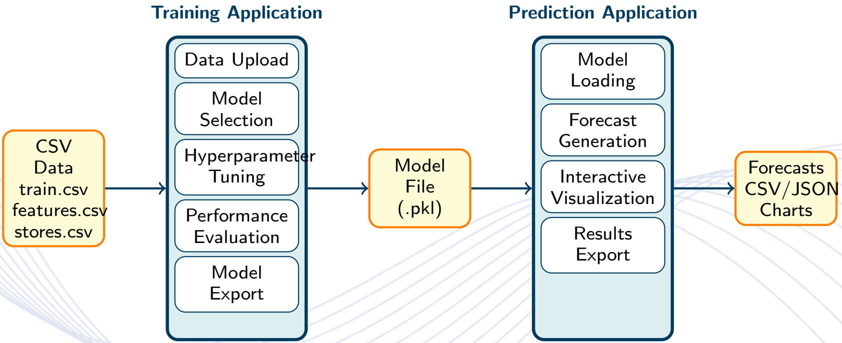
Walmart Sales Data Characteristics



System Architecture & Applications

Dual-Application Architecture I

Complete Forecasting Workflow



Dual-Application Architecture II

Complete Forecasting Workflow

Training Application

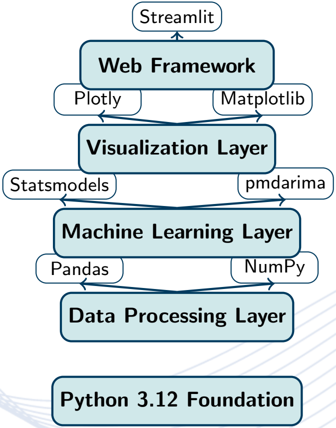
- Model development & validation
- Hyperparameter tuning
- Performance evaluation
- Model export capabilities

Prediction Application

- Production forecasting
- Interactive visualizations
- Real-time results
- Multiple export formats

Technology Stack Architecture I

System Overview and Component Structure



Models

Auto ARIMA

Exponential Smoothing

Holt-Winters

Technology Stack Details I

Core Technologies and Implementation

Core Technologies

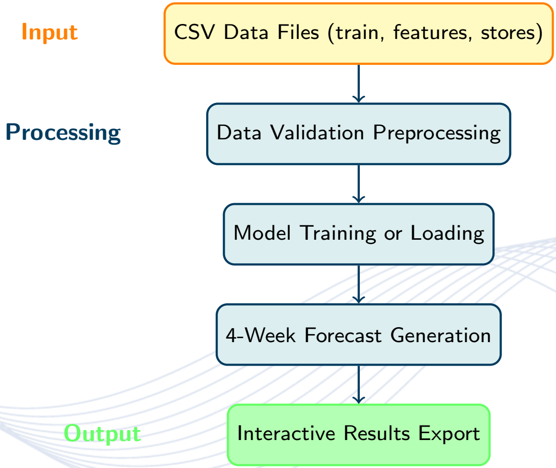
- **Python 3.12**: Exact version requirement
- **Streamlit**: Web application framework
- **Plotly**: Interactive visualizations
- **Pandas/NumPy**: Data processing

Forecasting Models

- **Auto ARIMA**: Automated parameter selection
- **Exponential Smoothing**: Holt-Winters method
- **Joblib**: Model serialization
- **Statsmodels/pmdarima**: Implementation

Data Pipeline I

From Raw Data to Forecasts



Data Pipeline II

From Raw Data to Forecasts

Input Processing

- **train.csv**: Historical sales data with store, date, weekly sales
- **features.csv**: External factors (temperature, fuel price, CPI, unemployment)
- **stores.csv**: Store metadata (type, size)

Output Generation

- **4-week forecasts**: Week-over-week sales changes
- **Interactive charts**: Color-coded visualizations
- **Export formats**: CSV, JSON for further analysis

Deployment Options I

Flexible Access Methods

Cloud Deployment

- Browser Access
- No Installation
- Automatic Updates
- Cross-Platform

VS

Local Installation

- Full Performance
- Offline Capability
- Data Privacy
- Large Datasets

Choose Based On:

Quick Start → Cloud
Production Use → Local
Sensitive Data → Local

Deployment Options II

Flexible Access Methods

Cloud Deployment

- Immediate browser access
- No installation required
- Automatic updates
- Cross-platform compatibility

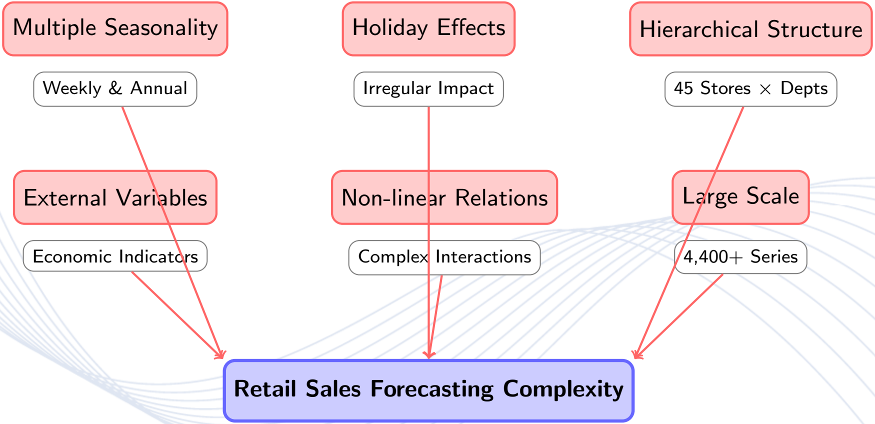
Local Installation

- Full performance control
- Offline capability
- Large dataset support
- Data privacy assurance

Technical Challenges & Solutions

Forecasting Challenges I

Complex Retail Sales Patterns



Forecasting Challenges II

Complex Retail Sales Patterns

Multiple Seasonal Patterns

- **Weekly seasonality**: Day-of-week effects
- **Annual seasonality**: Yearly consumption patterns
- **Holiday effects**: Irregular but significant impacts [MMH18]

Data Complexity

- **Hierarchical structure**: 45 stores $\times$ multiple departments
- **External variables**: Economic indicators integration
- **Non-linear relationships**: Traditional models limitations [PS17]

Implementation Challenges I

Technical and Deployment Issues

Cross-Platform Compatibility

- Python Dependencies
- Package Compatibility
- OS Variations

Performance Optimization

- Memory Management
- Processing Speed
- Interactive Response

Deployment Complexity

- Model Serialization
- Cloud Limitations
- User Experience

User Interface Design

- Intuitive Design
- Interactive Visualization
- Accessibility

Implementation Challenges II

Technical and Deployment Issues

Cross-Platform Compatibility

- Python version dependencies
- Package compatibility issues
- Operating system variations
- Model serialization consistency

Performance Optimization

- Memory management for large datasets
- Processing speed requirements
- Interactive response times
- Concurrent user support

Modeling Solutions I

Advanced Time Series Approaches

Auto ARIMA Implementation

- **Automated parameter selection**: Grid search across parameter space
- **Seasonal detection**: Automatic identification of patterns
- **AIC-based optimization**: Information criteria for model selection

Exponential Smoothing (Holt-Winters)

- **Triple smoothing**: Level, trend, and seasonal components
- **Flexible seasonality**: Additive or multiplicative options
- **Fast computation**: Suitable for real-time applications

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Modeling Solutions II

Advanced Time Series Approaches

Auto ARIMA

- Automated Parameter Selection
- Seasonal Pattern Detection
- AIC-based Optimization
- High Flexibility

VS

ETS (Holt-Winters)

- Triple Smoothing Components
- Flexible Seasonality Options
- Fast Computation
- Highly Interpretable

Performance Comparison

Auto ARIMA: More flexible, slower training
Exp. Smoothing: Faster, excellent for seasonality
Default Model: ETS(Holt-Winters) 3.58% WMAPE (Excellent)

Technical Solutions I

Web Application Development Strategy

Development Framework

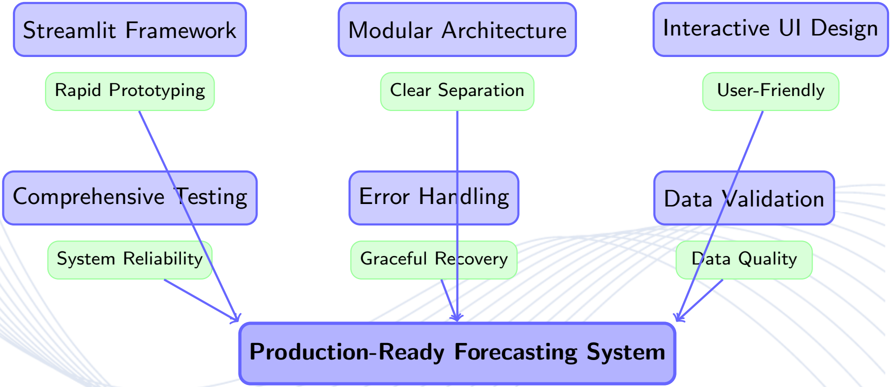
- **Streamlit**: Rapid prototyping capability
- **Modular architecture**: Separate training/prediction
- **Interactive UI**: User-friendly interface design
- **Real-time feedback**: Progress indicators and validation

Quality Assurance

- **Comprehensive testing**: Pytest validation suite
- **Error handling**: Graceful failure recovery
- **Data validation**: Schema and format checking

Technical Solutions II

Web Application Development Strategy



Results & Performance

Model Performance Results I

Quantitative Evaluation & Benchmarking

Model Performance

Default: 3.58% WMAE

Category: **EXCELLENT**

Horizon: 4 weeks

Business Impact

- High Confidence
- Reliable Predictions
- Actionable Insights

Our Model

↑ **Performance Categories**

less 5%	5-15%	more 15%
Excellent	Acceptable	Poor

Figure: Model Performance Comparison (WMAE Scores)

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Model Performance Results II

Quantitative Evaluation & Benchmarking

Performance Metrics

- **Default Model**: Holt-Winters
- **WMAE**: 3.58% (Excellent category)
- **Absolute Error**: $923.12 weekly
- **Forecast Horizon**: 4 weeks

Business Impact

- **95%+ Accuracy** for business planning
- **Reliable predictions** for inventory management
- **Seasonal patterns** effectively captured
- **Holiday effects** properly modeled

System Performance & Capabilities I

Technical Performance Metrics

Processing Performance

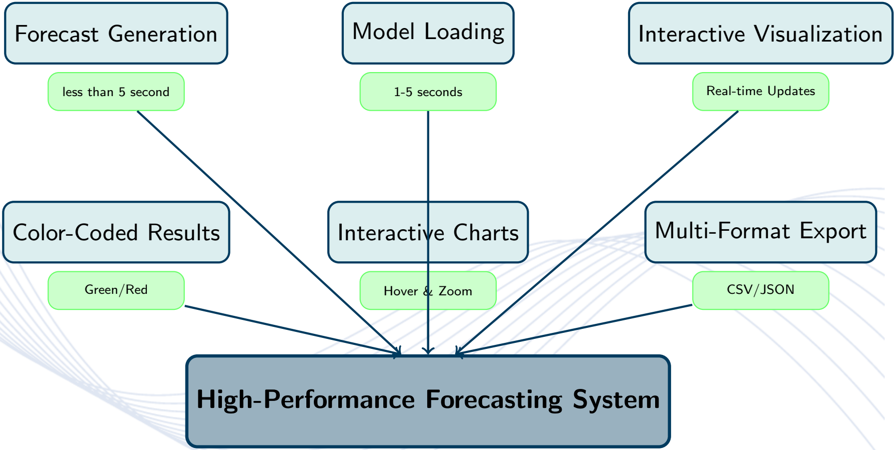
- **Model Loading**: Less than 5 seconds
- **Forecast Generation**: Less than 5 second
- **Visualization Rendering**: 1-5 seconds
- **Data Export**: Less than 10 second

Scalability

- **Cloud Users**: 50+ concurrent
- **Dataset Size**: Up to 200MB
- **Time Series**: 4,400+ supported
- **Memory Usage**: Optimized for web deployment

System Performance & Capabilities II

Technical Performance Metrics



System Performance & Capabilities III

Technical Performance Metrics

Figure: Overall System Performance

Key Achievements

End-to-End Workflow: From data upload to business insights in under 10 minutes, making sophisticated forecasting accessible to business users without technical expertise.

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Live System Demonstration I

Interactive Forecasting in Action

Demo Agenda

1. **Access** cloud application
2. **Load** default model
3. **Generate** 4-week forecast
4. **Interpret** results
5. **Export** data
6. **Show** training interface

What to Expect

- **Model Performance**: 3.58% WMAE display
- **Interactive Charts**: Color-coded forecasts
- **Business Insights**: Week-over-week changes
- **Export Options**: CSV/JSON downloads

Live System Demonstration II

Interactive Forecasting in Action

Live URLs

Prediction App: `walmart-sales-prediction-app-py.streamlit.app`
Training App: `walmart-sales-training-app-py.streamlit.app`

Backup Plan

Locals Streamlit Apps run if live demo encounters technical issues. ! :)

Results Interpretation I

Understanding Business Insights

Figure: Sample Forecast Output

Results Interpretation II

Understanding Business Insights

Key Insights

- **Green Bars**: Sales increase from previous week
- **Red Bars**: Sales decrease from previous week
- **Values**: Dollar amount of change
- **Trend**: Overall direction assessment

Business Actions

- **Positive Weeks**: Prepare inventory, schedule staff
- **Negative Weeks**: Optimize costs, plan promotions
- **Cumulative**: Overall month planning

Results Interpretation III

Understanding Business Insights

Important Note

Forecasts show **week-over-week changes**, not absolute sales values. This enables better understanding of sales momentum and trend direction.

Conclusions & Future Work

Key Contributions I

Research & Technical Achievements

Technical Contributions

1. **Integrated Web Platform**
   - First comprehensive web-based time series forecasting system
   - Unified training and prediction workflow
   - Cross-platform deployment (cloud + local)
2. **Advanced Model Integration**
   - Auto ARIMA + Holt-Winters implementation
   - Automated hyperparameter optimization
   - Performance-based model selection
3. **Interactive Business Intelligence**
   - Real-time forecast visualization
   - Color-coded trend indicators
   - Actionable business insights

Key Contributions II

Research & Technical Achievements

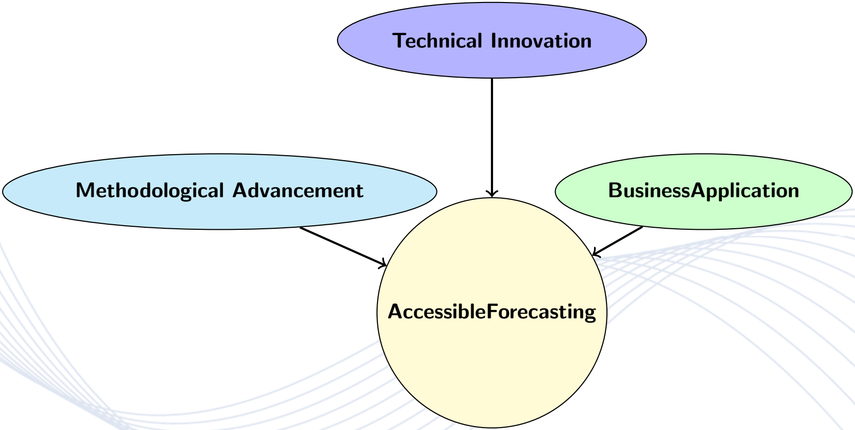


Figure: Contribution Areas

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Limitations & Lessons Learned I

Honest Assessment & Insights

Current Limitations

- **Temporal Scope**
  - Data: 2010-2012 (historical)
  - May not capture recent market changes
  - Limited long-term trend analysis
- **Forecast Horizon**
  - Fixed 4-week prediction window
  - No dynamic horizon adjustment
  - Limited seasonal cycle coverage
- **Model Scope**
  - ARIMA + Holt-Winters only
  - No ensemble methods
  - Limited external variable integration

Limitations & Lessons Learned II

Honest Assessment & Insights

Lessons Learned

- **Web Deployment Complexity**
  - Cross-platform compatibility critical
  - Model serialization challenges
  - User experience paramount
- **Business Focus Essential**
  - Technical accuracy $\neq$ business value
  - Interpretation matters more than precision
  - Accessibility drives adoption
- **Iterative Development**
  - Continuous testing essential
  - User feedback invaluable
  - Performance optimization ongoing

Limitations & Lessons Learned III

Honest Assessment & Insights

Key Insight

The most sophisticated algorithm is useless if it's not accessible to the people who need to make decisions based on its output.

Future Work & Enhancements I

Roadmap for Continued Development

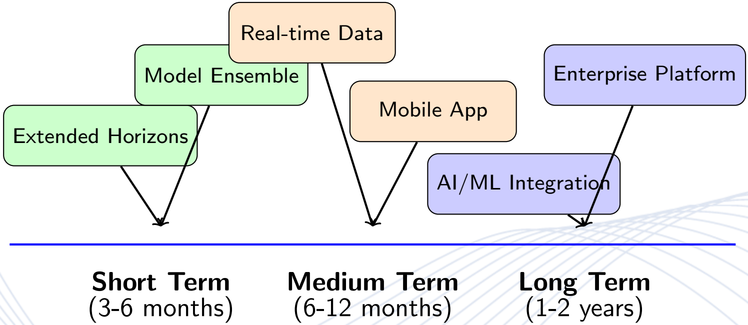


Figure: Development Roadmap

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Future Work & Enhancements II

Roadmap for Continued Development

Short Term (3-6 months)

- **Extended Horizons**: 8-12 week forecasts
- **Ensemble Methods**: Model combination
- **Performance Tuning**: Speed optimization
- **Additional Models**: Prophet, LSTM

Medium Term (6-12 months)

- **Real-time Data**: API integration
- **Mobile Application**: iOS/Android apps
- **Advanced Analytics**: Confidence intervals
- **User Management**: Multi-user support

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Future Work & Enhancements III

Roadmap for Continued Development

Long Term (1-2 years)

- **AI Integration**: AutoML capabilities
- **Enterprise Platform**: Commercial deployment
- **Industry Expansion**: Beyond retail
- **Research Platform**: Academic collaboration

University of Applied Sciences
HOCHSCHULE EMDEN·LEER

Final Thoughts I

Impact & Significance

Project Impact

- **Academic Contribution**
  - Demonstrates practical time series deployment
  - Bridges theory-practice gap
  - Provides open-source foundation
- **Business Value**
  - Accessible forecasting for SMEs
  - Rapid prototype development
  - Cost-effective solution
- **Technical Innovation**
  - Web-based ML deployment patterns
  - Cross-platform compatibility solutions
  - Interactive visualization best practices

Final Thoughts II

Impact & Significance

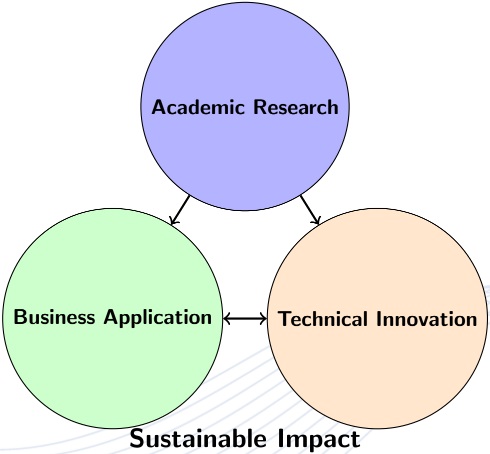


Figure: Multi-dimensional Impact

Thank You

for your attention

Sources

Quellen I

[FG19] Robert Fildes and Paul Goodwin. "Retail forecasting: Research and practice". In: *International Journal of Forecasting* 35.1 (2019), pp. 1–9. DOI: 10.1016/j.ijforecast.2018.09.008. URL: https://www.sciencedirect.com/science/article/abs/pii/S016920701930192X.

[Loy17] Joshua D. Loyal. "The Walmart Sales Project". In: *Unpublished Manuscript* (2017). URL: https://joshloyal.github.io/assets/pdf/forecasting_intro.pdf.

[Man+22] S. Mane et al. "Comparative Analysis of ML Algorithms & Stream Lit Web Application". In: *2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*. IEEE. 2022, pp. 1–6. DOI: 10.1109/ICDCECE53908.2022.9725496.

Quellen II

[MMH18] Tucker S. McElroy, Brian C. Monsell, and Rebecca J. Hutchinson. "Modeling of Holiday Effects and Seasonality in Daily Time Series". In: *U.S. Census Bureau Working Papers* (2018). URL: https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/rrs2018-01.pdf.

[PS17] James J. Pao and Danielle S. Sullivan. "Time Series Sales Forecasting". In: *CS229: Machine Learning Final Projects* (2017). URL: https://cs229.stanford.edu/proj2017/final-reports/5244336.pdf.

Quellen III

[Sur+23] Aaditya Surya et al. “Enhanced Breast Cancer Tumor Classification using MobileNetV2: A Detailed Exploration on Image Intensity, Error Mitigation, and Streamlit-driven Real-time Deployment”. In: *arXiv preprint arXiv:2312.03020* (2023).

[Zha21] Jiayuan Zhang. “Sales Prediction of Walmart Based on Regression Models”. In: *Proceedings of the 2021 International Conference on Computers, Information Processing and Advanced Education*. 2021, pp. 294–298. DOI: 10.1145/3456887.3459308. URL: https://www.atlantis-press.com/article/125994715.pdf.