📋 EXECUTIVE SUMMARY - COMPARATIVE MODEL ANALYSIS

🎯 PRIMARY OBJECTIVE

Develop a comprehensive comparative analysis framework between Neural Networks and Linear Regression for electrical load forecasting, evaluating performance across different temporal horizons (24 hours vs 15 days) with risk analysis through Cumulative Distribution Functions (CDFs).

🔧 IMPLEMENTED METHODOLOGY

1. Automated Data Preparation

- **Automatic feature detection:** 39+ variables identified automatically
- **Cyclical feature engineering:** Sin/cos transformations for temporal variables (hour, day, month)
- **Stratified splitting:** 70% training, 20% validation, 10% test
- **Normalization:** StandardScaler applied to neural networks only

2. Model Architectures

Feedforward Neural Networks:

- Architecture: 39 → 128 → 64 → 32 → 1
- Optimizer: ADAM (lr=0.001)
- Training: 100 epochs with early stopping
- Separate models for 24h and 15-day horizons

Linear Regression:

- Ordinary least squares method
- Same 39 features as neural networks
- Independent models per temporal horizon

3. Multi-Horizon Evaluation

- **Short horizon:** 24 hours (`load_24h`)
- **Long horizon:** 15 days (`load_360h`)
- **Metrics:** RMSE, MAE, R^2

📊 ADVANCED VISUAL ANALYSIS (5 Visualizations)

1. Cumulative Distribution Functions (CDFs)

- **Purpose:** Risk analysis and error distribution assessment
- **Insight:** Enables evaluation of extreme error probability

2. Comparative Scatter Plots (2×2 Panel)

- **Predictions vs actual values** for all models
- **R^2 annotated** on each plot for direct evaluation

3. Comparative RMSE Analysis

- **Comparative bars** by model and temporal horizon
- **Visual identification** of best performance

4. Training Evolution + Percentiles

- **Learning curves** for neural networks
- **Error percentile analysis** (P50, P75, P90, P95, P99)

5. Integrated Panel

- **Combined view** of CDFs and scatter plots
- **Comprehensive analysis** in single visualization

🎯 SCIENTIFIC CONTRIBUTION

Methodological Innovation:

1. **Systematic multi-horizon analysis** (24h vs 15 days)
2. **CDFs for risk analysis** in load forecasting
3. **Automated model comparison framework**
4. **Robust cyclical feature engineering**

Statistical Rigor:

- **Multiple evaluation metrics** (RMSE, MAE, R^2)
- **Percentile analysis** for risk management
- **Appropriate temporal cross-validation**
- **Differentiated normalization** by model type

💡 FOR EXPLAINING TO YOUR PROFESSOR

"I developed a comparative analysis system that automatically evaluates two machine learning approaches (neural networks vs linear regression) for electrical load forecasting across different temporal horizons. The key innovation is the use of Cumulative Distribution Functions (CDFs) for risk analysis, enabling quantification of extreme error probability in predictions. The framework automatically generates 5 advanced visualizations that reveal not only which model performs better, but also how much risk each approach carries across different temporal horizons."

Added Value:

- **Reproducibility:** Fully automated code
- **Scalability:** Framework extensible to other models
- **Interpretability:** Professional-grade visualizations
- **Applicability:** Methodology transferable to other domains

Key Research Questions Addressed:

1. **Model Performance:** Which approach (NN vs LR) performs better?
2. **Temporal Degradation:** How does accuracy change with prediction horizon?
3. **Risk Assessment:** What's the probability of large prediction errors?
4. **Practical Implementation:** Which model is more reliable for operational use?

Technical Innovation:

- **Automated pipeline** from raw data to publication-ready plots
- **Multi-dimensional comparison** across models, metrics, and time horizons
- **Risk-aware evaluation** beyond traditional accuracy metrics
- **Professional visualization framework** for scientific communication