Machine Learning & Statistical Prediction Models to forecast Supplier Performance to Evaluate Supplier Risk

(Tech Stack: Python, Pandas, Matplotlib, AppScript)

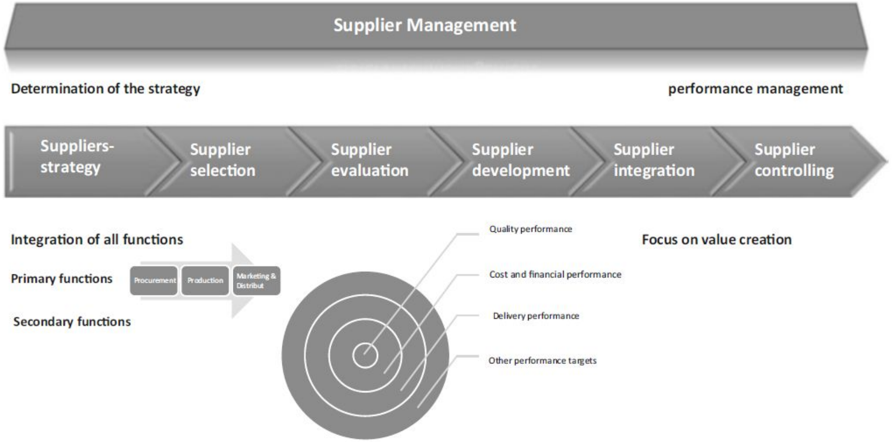
- Heramb Joshi

Table of Contents

Introduction

Background & Motivation

- Increase in complexity in global Supply Chains
- Importance of Supplier Risk Analysis to reduce supply chain disruptions
- Measuring & forecasting supplier performance
- Challenges faced - Missing and/or inaccurate data



Research Goals

- Impute missing supplier performance data
- Apply advanced statistical & ML models to test the hypothesis
- Evaluate effectiveness of models in improving risk prediction accuracy

Research Questions

- How can we identify a critical (low performing) supplier with forecast error when evaluating the supplier risk to avoid supply chain disruption?
- How does predictive analytics improve supplier performance evaluation with missing or inaccurate data?
- What statistical models are most effective for forecasting missing data points in supplier performance matrix?

Proceedings

Research Methodology

- Quantitative methodology used collecting secondary data from company TSM Supply Bridge
- Experimental research was conducted on supplier performance data to test hypothesis (research questions)

Data Overview

- Time series dataset consists of month-by-month supplier performance data watchlists:
  - Finance - 99 Suppliers data of 21 months
  - Sustainability - 189 Suppliers data of 16 months
  - Quality - 881 Suppliers data for 14 months
  - OTD - 881 Suppliers data for 14 months

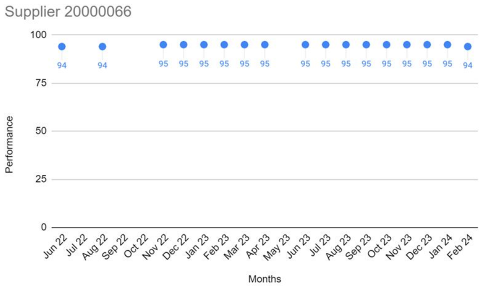
Suppliers	Jun 22	Jul 22	Aug 22
20000021	84	84	84
20000025	76.36666667		77.484375
20000066	94	94	94
20000154	76.36666667	83.68333333	91
20000185	74	81.5	89
20000207	76.36666667	73.58611111	70.80555556
20000271	78	78	78
20000614	76.36666667	75.00277778	73.63888889
20000669	100	100	100
20000676	76	76	76
20000809	53	53	53
20000831	97	97	97
20000921	61	61	61

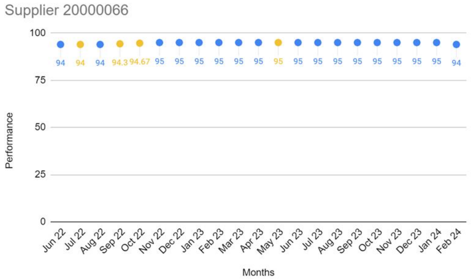
- In this research -
  - Only Finance KPI used for analysis as larger monthly dataset available
  - Measured in 0 (low performance) - 100 (high performance) points
  - Additional KPIs from dataset like PPM, Purchasing Volume & Failure costs not analyzed

Proceedings

Missing Data Challenge

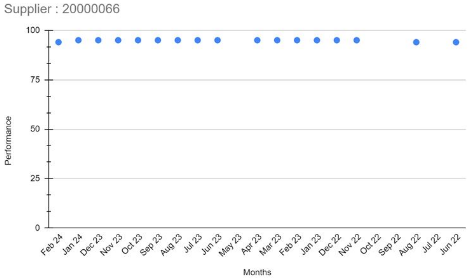
- Missing values in KPI weaken supplier performance assessments
- Imputation methods used
  - Linear Interpolation,
  - K Nearest Neighbor,
  - Forward/ Backward Fill,
  - Mean/Median/Mode

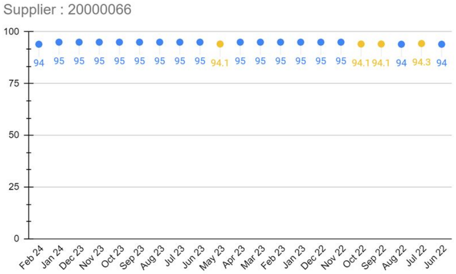




Before & After Linear Interpolation

Before & After KNN





Proceedings

Prediction Models

- Monte Carlo Simulation: Simulates thousands of scenarios to forecast future performance, providing a probabilistic range of outcomes.
- Markov Chain: Models supplier performance as a sequence of "states" (e.g., low, medium, high performance) and calculates the likelihood of transitions between these states over time.
- LSTM Networks: A type of deep learning model designed for sequential data (time series). LSTM accounts for long-term dependencies and patterns in performance data to predict future performance.

Model Training

- Used 80% of the data for training and 20% for validation.
- Accuracy and forecast error were assessed for each model using mean absolute error (MAE)
- Employed cross-validation to reduce overfitting and ensure model generalization

Results

Key Results

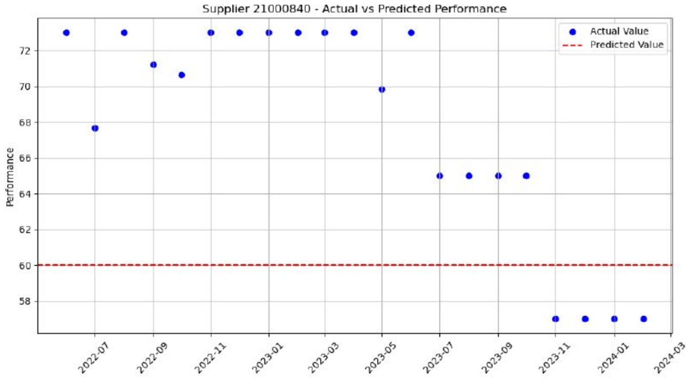
- The LSTM model's strong performance highlights the importance of capturing long-term dependencies.
- Monte Carlo Simulation provides a probabilistic approach to risk assessment and accurate prediction in the case of irregular data trends.
- Markov Chain analysis offers insights into the likelihood of state transitions with focus on the present state more than historical data.

Effectiveness of Predictive Models based on the Different Data Patterns

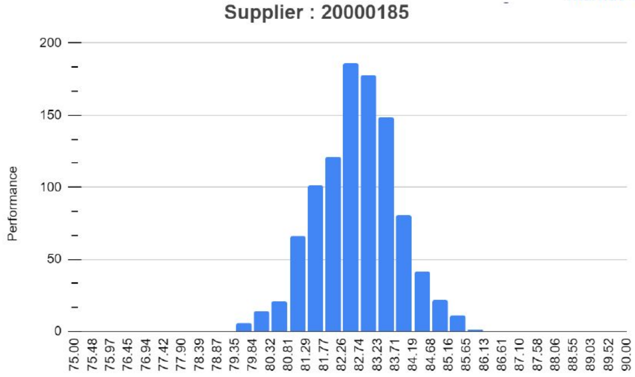
Data Pattern	Markov Chains Model	LSTM Model	Monte Carlo Simulation Model
Stationary Patterns	Strong	Moderate	Moderate
Linear Trends	Weak	Strong	Weak
Exponential Trends	Weak	Strong	Weak
Seasonal Patterns	Moderate	Strong	Moderate
Cyclical Patterns	Moderate	Strong	Moderate
Irregular Patterns	Weak	Moderate	Strong
Non-Stationary Patterns	Weak	Strong	Moderate
Complex Temporal Dependencies	Weak	Strong	Moderate

Understanding the pattern in any given KPI will help in making the decision about the approach to be used.
This makes the models relative and can be applied to any KPI.

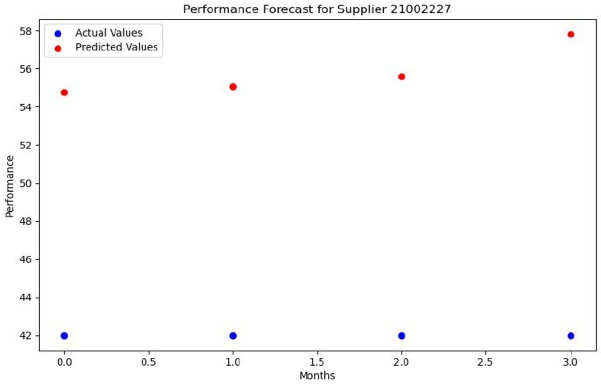
Key Results



Markov Chain Results



Monte Carlo Simulation Result



LSTM Results

Key Results

Comparison of Models

- Top 5 Suppliers with highest forecast error →

Table 1. Comparison of Imputation Methods

Imputation Method	Use Case	Advantages	Drawbacks	Inference and Best for
Linear Interpolation	Short gaps in time-series data. - KPIs with a clear, linear trend over time.	Simple, preserves trend direction	Not suitable for non-linear trends. Fails when large gaps	Small gaps (1%-19% missing data) with an expected linear trend
Forward/ Backward Fill	Short term predictions in time-series	Effective for small and frequent gaps	Not suitable for long gaps Error if last known or future values non representative	Time-series data with short gaps, where past/ future trends are more predictable
KNN Imputation	Data with complex relationship, where there is significant missing data	No assumption about data distribution	Computationally heavy for large datasets and sensitive to distance metric	Complex datasets with moderate or high missing values 20% - 50%

Table 3. Comparison of Predictive Model's Performance

Supplier	% Missing Data	AV	Predicted Values			Forecast Error			Minimum FV
			LSTM	Monte Carlo Sim.	MC	LSTM	Monte Carlo Sim.	MC	
21000840	19.05	57	63.5778236 4	68.26035	60	7.143769264	-11.26035	-3	Markov Chain
20001565	23.81	79	95.2543029 8	89.82114	85	8.235124588	-10.82114	-6	Markov Chain
20001367	19.05	82	90.9686279 3	80.20420	85	9.488697052	1.7958	-3	Monte Carlo
21002227	19.05	42	51.2977027 9	73.03220	60	10.81674385	-31.0322	-18	LSTM
20000185	19.05	76	89.2172699 0	80.43010	85	13.91151428	-4.4301	-9	Monte Carlo

Where -
AV = Actual Values
FV = Forecast Value
FF = Forward Flip
LI = Linear Interpolation
MC = Markov Chain Approach
Sim = Simulation

Key Results

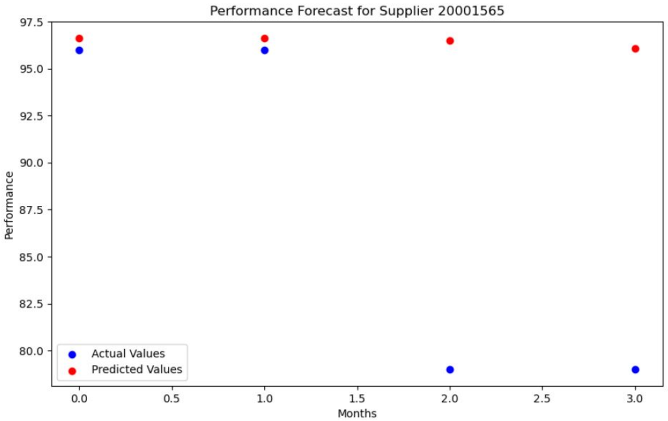
Supplier Risk Analysis

- Identifying critical Supplier based on forecast error.
  - Concluded that more the forecast error more the supplier is risky as the supplier is unreliable
- The LSTM model outperformed other models for supplier 20001565 with a MAE of 8.235, indicating strong ability to capture long-term dependencies in supplier performance data.
- Impact of Predictive Analytics: Advanced data imputation and forecasting techniques significantly improved risk prediction

Conclusion

- This project,
  - measured and forecasted supplier performance for supplier risk evaluation
  - addressed missing data by data imputation methods

enabling the supply chain managers for decision making



References

Dust, R., 2019. Total Supplier Management. München: Hanser Verlag .

Dust, R., Goldschmit, J. & Gürtler, B., 2011. Total Supplier Risk Monitoring. Berlin: Datenqualität als zwingende Grundlage einer effektiven Lieferantenbewertung.

Fischer, M., 2020. Markov chains and Processes. s.l.:Saint-Étienne School of Mines.

Glasserman, P., 2003. Monte Carlo Methods In Financial Engineer. New York: Springer.

Helmold, M., 2023. Innovative Supplier Management. Berlin: Springer.

Hochreiter, S., 1997. Long Short Term Memory. Neural Computation.

Kersten, W. (., Blecker, T. (. & Ringle, C. M. (., 2020. Data Science and Innovation in Supply Chain Management: How Data Transforms the Value. Hamburg, No. 29, ISBN 978-3-7531-2346-2, epubli GmbH, Berlin, https://doi.org/10.15480/882.3100.

Kleemann, F. a. G. A., 2020. Einkauf 4.0. Digitale Transformation der Beschaffung. 2. Auflage.. Wiesbaden: Springer Gabler

Li, H. & Li, X., 2023. A Case Study on Analyzing the Similarity Between Two Similar Stocks: Building the CrC-LSTM Model. IEEE International Conference on e-Business Engineering (ICEBE) , pp. 189 -191.

Mohamed Noor, N. &. A. M. M. A. B. &. Y. A. S. &. R. N., 2014. Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set. Materials Science Forum. 803. 278-281. 10.4028/www.scientific.net/MSF.803.278.