Dablio, Gleendon T.

MERCADO LIBRE MONTHLY SALES

Executive Summary:
This report provides an analytical review of Mercado Libre's sales performance using daily sales data from January 2019 to May 2020. Mercado Libre, the leading e-commerce and fintech platform in Latin America, offers invaluable insights into regional market dynamics through its operations in 18 countries. The objective of this study was to uncover patterns and predict future sales trends to inform business strategies.

Introduction:
Mercado Libre (NASDAQ: MELI), founded in 1999 by Marcos Galperin in Argentina, has evolved into Latin America's most prominent e-commerce and technology company. Operating in 18 countries, it serves as the region's largest online marketplace and fintech platform, often referred to as the "Amazon of Latin America."

Mercado Libre's dataset contains daily sales performance metrics from January 2019 to May 2020, representing a critical period of the company's growth. Understanding the patterns, trends, and anomalies in this time series data is crucial for strategic decision-making and operational optimization.

Statement of the Problem:
The primary challenge is to analyze the temporal patterns and variations in Mercado Libre's daily sales performance to identify growth trends, seasonality, and potential influencing factors that could impact business strategies and forecasting models.

Objectives of the Analysis:
- To group daily sales into monthly
- To identify trends and seasonality
- To create a model that forecast 6 months later
- To validate different types of model for best model fit

Data Description:
The Data is consist of 2 columns which is date and Daily Sales it has 500 rows if converted to month it will have 17 months from january 2019 to may of 2020

Methodology:

Model Selection:

- Selected a diverse set of machine learning and deep learning models suitable for time-series forecasting:
    - Deep Neural Networks (DNN)
    - Convolutional Neural Networks (CNN)
    - Long Short-Term Memory Networks (LSTM)

- Recurrent Neural Networks (RNN)

Cross-Validation:

- Implemented cross-validation techniques to evaluate the models' performance across different subsets of the dataset, ensuring robust assessments of their predictive accuracy and generalizability.

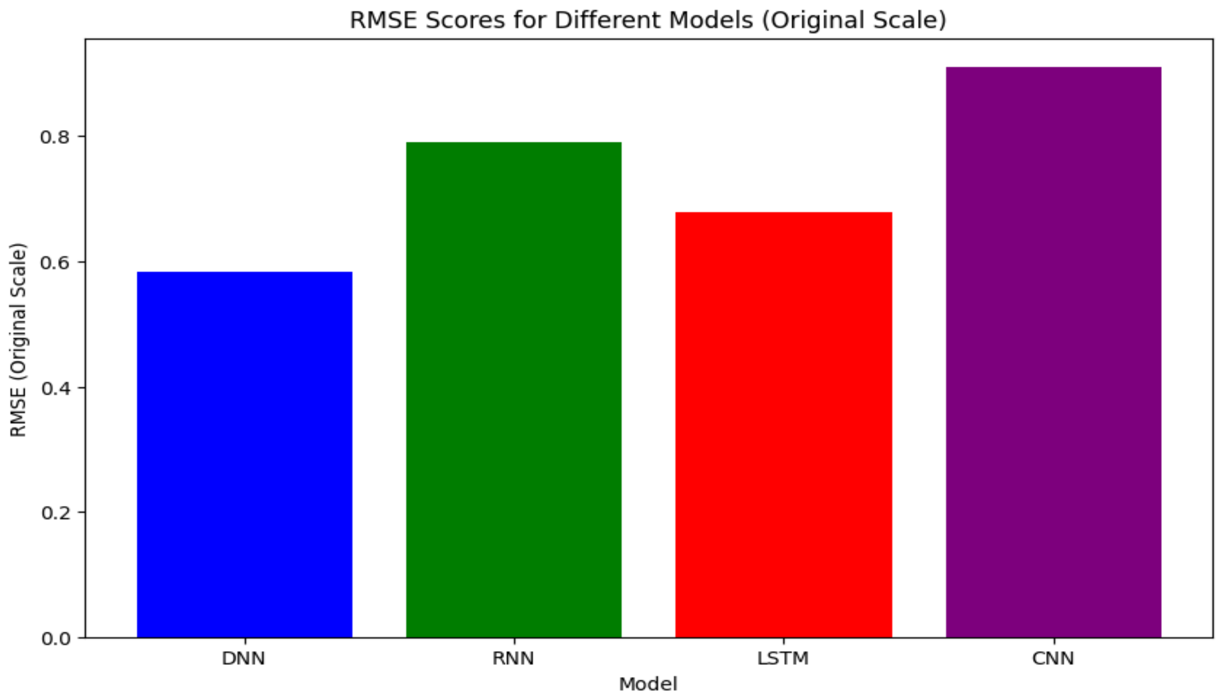
Error Metrics:

- Used Root Mean Square Error (RMSE) as the primary metric for evaluating and comparing model performance.
- RMSE provided insights into the magnitude of prediction errors for each model.

Model Performance Comparison:

- Conducted a detailed comparison of model performances:
    - Assessed the ability of RNNs and LSTMs to handle sequential patterns and long-term dependencies in the time-series data.
    - Evaluated CNNs for their capacity to detect local patterns and trends in sales data.
    - Benchmarked the simplicity and efficiency of DNNs against more complex architectures.

Cross-validation of 4 models




```
RMSE Scores (Original Scale):
DNN: 0.5829792602042715
RNN: 0.7897673826424766
LSTM: 0.6779174906311616
CNN: 0.9109443604501736
```

Model Comparison:

- The bar chart compares the RMSE (Root Mean Square Error) scores of four models: DNN (Deep Neural Network), RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory), and CNN (Convolutional Neural Network). Lower RMSE scores indicate better performance.

DNN Superiority:

- Among the models, the DNN has the lowest RMSE, suggesting it performs better in minimizing prediction errors compared to the other models.

Scale of Errors:

- The RMSE values range between approximately 0.6 and 0.9, measured on the original scale of the data. The CNN exhibits the highest error, while the DNN has the smallest.

Model Complexity vs Performance:

- The performance does not consistently align with the expected complexity of the models. For instance, while RNN and LSTM are designed for sequential data, their RMSE is not significantly better than the simpler DNN. The CNN, often associated with image data, underperforms in this comparison.

Implications for Model Selection:

- Given that DNN shows the best RMSE score, it may be the preferred model for this task. However, other factors like training time, interpretability, and the nature of the dataset should also influence the selection.

Data Representation:

- The results suggest that the dataset may not favor sequential or spatial data modeling, as seen from the higher RMSE scores of RNN, LSTM, and CNN. The DNN might have effectively captured the patterns due to simpler underlying data structures.

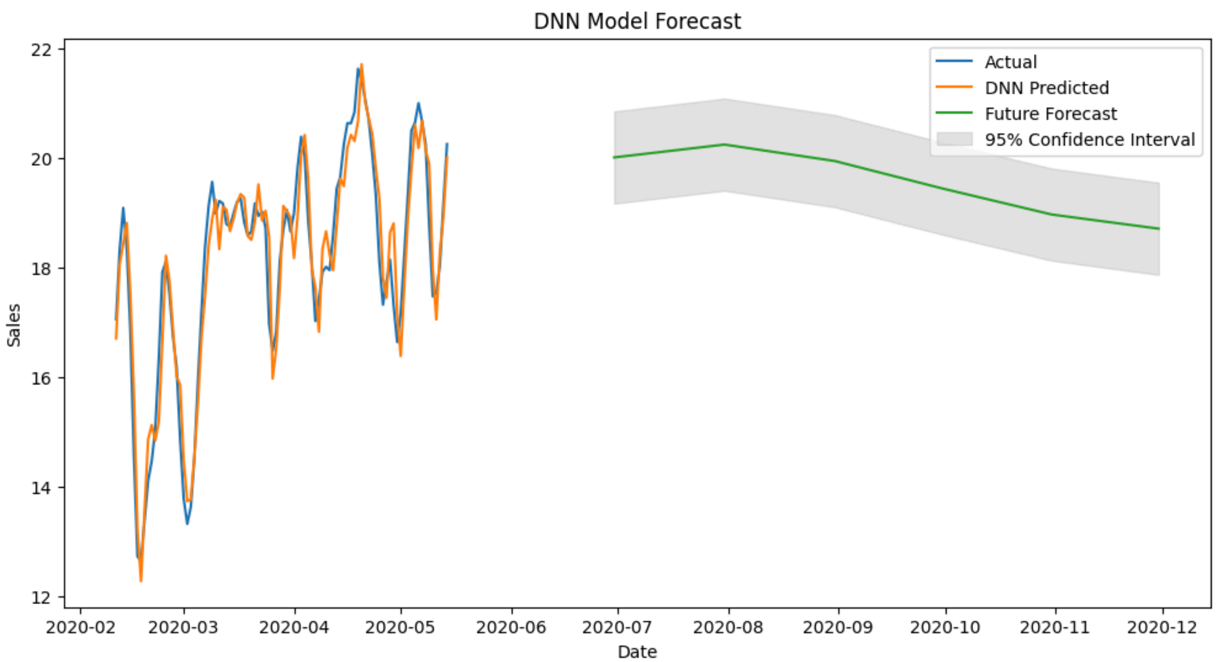
Historical Performance:

- If similar tasks historically show superior performance for DNNs, this further validates the model's suitability for the current dataset. Conversely, the poor performance of CNN could indicate a mismatch between the model's architecture and the data's features.

Model Fit:

- The **best model** in this comparison is the **DNN (Deep Neural Network)** because it has the lowest RMSE score, indicating it performs best at minimizing prediction errors. Lower RMSE suggests the model's predictions are closer to the actual values, making it the most accurate model among the four options evaluated.

Future Forecast:

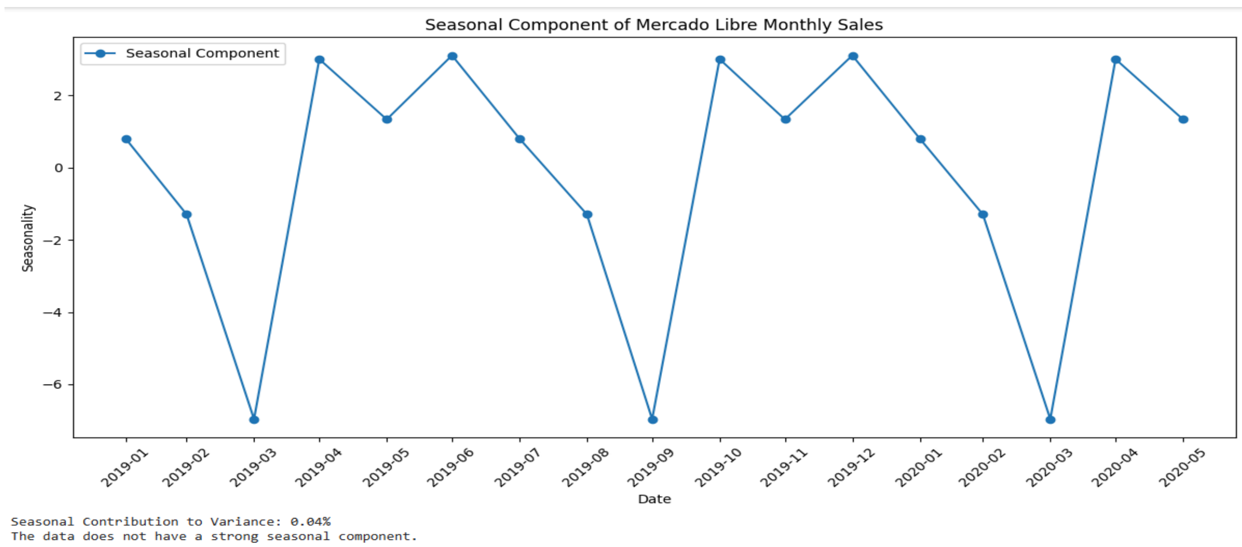


The forecast is **relatively stable** but shows a **gradual decline** in sales.
The uncertainty grows as the forecast moves further into the future, as seen by the expanding width of the confidence interval.
This indicates that while the model is confident in short-term predictions, long-term predictions are less certain due to possible variations in the data.

Uncertainty:

- The confidence interval widens as the forecast moves further into the future.
- This reflects the **inherent difficulty of making accurate long-term predictions**, as the influence of unknown variables and noise grows over time.

Seasonality:



- In other words, sales patterns are primarily driven by other factors like trends or noise rather than seasonal effects.

Overall Trend:

- The sales data does not show a strong upward or downward trajectory over the observed period (2019–2020).
- There may be slight dips in certain months, but the general trend appears relatively stable.

Conclusion:

The analysis demonstrates that while multiple forecasting models were evaluated, the **DNN model** emerged as the most effective due to its lower RMSE and stronger prediction accuracy. The forecast highlights a **gradual decline in sales** over time, with increasing uncertainty in long-term predictions as indicated by the widening confidence intervals.

Seasonality contributes minimally to sales variance (0.04%), suggesting that sales patterns are primarily influenced by trends and irregular factors rather than recurring seasonal effects. The overall trend indicates a **stable but slightly declining sales performance**, with no significant growth observed during the period analyzed.

Given these findings, the focus should shift toward understanding and addressing non-seasonal drivers of sales, such as market conditions, external events, and product demand. While the DNN model is well-suited for short-term forecasting, its long-term predictions should be used cautiously and in conjunction with scenario planning.

Recommendation:

- Focus on addressing non-seasonal factors such as market demand and competitive pressures to counteract the observed sales decline.
- Optimize short-term forecasting with DNN and explore hybrid models to enhance long-term prediction reliability.
- Conduct periodic reviews of the forecasting model to adapt to emerging trends and ensure robust decision-making.

company link:
https://global-selling.mercadolibre.com/landing/about

dataset source:
https://github.com/Dsmannight/MercadoLibre_forecast/blob/main/Resources/mercado_daily_revenue.csv