Project Title: Predicting Monthly Total Revenue for Business

Topic

Predicting Monthly Total Revenue for Business Using Machine Learning and Deep Learning Techniques

Business Problem

The project aims to predict the next month's total revenue using historical billing data processed and any manual adjustments booked to the general ledger. This prediction will help in financial planning, resource allocation, and strategic decision-making to enhance profitability and operational efficiency.

Datasets

The data is sourced from various billing systems within the company and stored in a Teradata database. They are all processed and interfaced to the ERP system. Later, processed billing and adjustment data are replicated to the Teradata analytics solution DB. Dataengineers further process the data and produce one fact table. The scope of this project is limited to billing data. So the dataset only contains billing information for 2 years.

The datasets include:

- **report_month:** accounting period in date when the bills are reported
- **billing_year:** accounting year when the bills are reported
- **billing_month:** accounting month when the bills are reported
- **billing_cycle:** accounting day (1 - 31) when the bills are reported
- **billing_day:** accounting work day (excluding weekends) when the bills are reported
- **billing_amount :** billed amount booked in General Ledger

Methods

The analysis methods for this project include:

1. **Data Extraction and Preprocessing:** Using SQL to extract and preprocess data from the Teradata database.
2. **Feature Engineering:** Creating relevant features from the raw data to enhance model performance.
3. **Incremental Training:** Developing a machine learning model (e.g., ARIMA, LSTM) that can be incrementally trained each month as new data becomes available.
4. **Model Evaluation and Validation:** Using metrics like MAE, RMSE, and MAPE to evaluate the model's performance.

Ethical Considerations

Potential ethical concerns include:

- **Data Privacy:** Ensuring that customer data is anonymized and secure to prevent any privacy breaches.
- **Bias and Fairness:** Avoiding bias in the model that could disproportionately affect certain customer segments or business areas.
- **Transparency:** Making sure the model's predictions are interpretable and explainable to stakeholders.

Challenges/Issues

Some challenges and issues that might be faced include:

- **Data Quality:** Ensuring the data extracted from different systems is clean, consistent, and accurate.
- **Model Adaptation:** Continuously adapting the model to handle changes in business processes or external factors that may impact revenue.
- **Scalability:** Efficiently managing and processing large volumes of data as the dataset grows over time.

References

Crone, S. F., Hibon, M., & Nikolopoulos, K. (2011). Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. *International Journal of Forecasting, 27*(3), 635–660. (https://www.sciencedirect.com/science/article/abs/pii/S0169207011000616?via%3Dihub)

Lazzeri, F. (2021). *Machine Learning for Time Series Forecasting with Python*. Wiley. (https://github.com/PacktPublishing/Machine-Learning-for-Time-Series-with-Python)

Lim, B., & Zohren, S. (2021). Time series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 379*(2194). (https://royalsocietypublishing.org/doi/10.1098/rsta.2020.0209)