Midterm Report

Brian Cervantes Alvarez
February 6, 2024
ST565 Time Series

Time Series Analysis on Enrollment Data

My Research Interest

One of my current research interests lies in exploring the relationship between student retention and financial aid, among various other factors. The paper for this report took my attention given the similar nature of student enrollment data, which aligns to the current dissertation that I am exploring.

Journal's Research Question

The paper by Urbano B. Patayon and Renato V. Crisostomo aims to forecast future student enrollment numbers at Jose Rizal Memorial State University – Tampilisan Campus, focusing on addressing the potential issues of resource allocation in response to enrollment changes. Their research question was: How can forecasting models predict future enrollment numbers, aiding university administrators in making informed decisions regarding infrastructure and human resource needs?

Tools & Models Considered

The authors utilized the Prophet forecasting model, an open-source tool developed by Facebook for time series forecasting that can handle daily, weekly, and yearly seasonality along with holiday effects. Prophet is known for its robustness to missing data, trend shifts, and outliers, making it a popular choice for various forecasting scenarios.

Methodology & Implementation

The implementation involved using historical enrollment data from 2000 to 2022, without explicitly detailing any pre-processing steps, exploratory data analysis, or validation of the model assumptions. The authors used Root Mean Square Error and the Coefficient of Determination, (R_2), as metrics to evaluate the model's performance.

Overall Impression

While the findings are potentially useful for administrative planning, the study clearly falls short on several methodological and analytically issues. In the next page, I highlight all the research red flags that I found while reading the paper. One infuriating point that I want to get across is the lack of grammar and proof reading. After a ten minute author research, it is evident that the authors have published other papers with similar shortcomings. Using my lens as a data scientist, these papers have to be republished given what I am going to say next.

Research Red Flags

Preliminary Data Analysis Deficiencies

Lack of Exploratory Data Analysis: Neglecting EDA and statistical testing (e.g. stationarity, seasonality, autocorrelation) misses crucial insights and fails to address data distribution and inherent characteristics, essential for model selection and validation.

Insufficient Statistical Analysis: Jumping directly to forecasting without preliminary analyses overlooks the characteristics of the data. This may have resulted in using an the incorrect forecasting tool such as Facebook's Prophet model. This is poor research practice to jump into using a model without analyzing the structure of the data.

Model Selection and Validation Shortcomings

Lack of Model Assumption Checks: Failing to validate the assumptions of the forecasting model may lead to inaccurate forecasts. There was no clear statement of the assumptions, which is poor research practice.

Inadequate Validation of Forecasting Results: Limited discussion on cross-validation techniques or error analysis restricts understanding of the model's predictive capabilities.

Overreliance on a Single Model: Solely using Prophet without comparing other methodologies limits evaluation of the most effective forecasting tool. No model comparison makes it difficult to justify their findings.

Analysis and Interpretation of Results

Limited Explanation of Findings: The authors provided numeric outcomes without detailed interpretation of trends or fluctuations in enrollment numbers. For example, the authors mentioned that Facebook's Prophet model is the sum of 4 times series models, which is given exactly as they stated:

$$y(t) = g(t) + s(t) + h(t) + e_t$$

where:

$g(t) =$ growth function
$s(t) =$ seasonality function
$h(t) =$ holiday/event function
$\varepsilon =$ error

However, they did not end up elaborating their final model in the context of this formula. It would have been a wonderful addition to see the final model from Prophet, even in a pseudo-manner.

Generalization of Results: They displayed a lack of discussion on the applicability of findings to other contexts such as the impact of external factors on enrollment trends.

Data Quality and Ethical Issues

Data Quality and Completeness: The study does not fully address issues like missing data, data entry errors, or the impact of external factors on model accuracy. More importantly, how did they even clean their data?

Ethical and Privacy Considerations: Neglects potential ethical and privacy considerations in using student enrollment data.

What I Recommend to the Publishers of this Paper

Given the current analysis, I recommend a "Revise and Resubmit" decision for the paper. The Urbano B. Patayon and Renato V. Crisostomo should be encouraged to the folowing:

- Conduct thorough exploratory data analysis and statistical testing to validate the assumptions of the forecasting model.
- Compare the Prophet model's performance with other forecasting approaches to establish its relative efficacy.
- Address the methodological issues raised, particularly the lack of detailed data analysis and discussion on model assumptions and limitations.

Improving on these aspects could significantly enhance the paper's contribution to the field of educational administration.

Reference

Patayon, U. B., & Crisostomo, R. V. (2022). Time Series Analysis on Enrolment Data: A case in a State University in Zamboanga del Norte, Philippines. 2022 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 13-18. https://doi.org/10.1109/ICACSIS56558.2022.9923436