GST REVENUE PREDICTION USING MACHINE LEARNING

- Ahaan Choksi
- Christopher George
- Tanmay Chowdhary

ABOUT GST

- The Goods and Service Tax is an indirect tax imposed on the supply of goods and services
- It was introduced on 1st July 2017, GST replacing multiple indirect taxes like VAT, service tax, and excise duty.
- It is divided into CGST, SGST, IGST (Central, State, and Integrated GST for interstate transactions)
- GST contributes over 60% of India’s indirect tax revenue.
- It is a key pillar of government finances.

Forecasting GST revenue plays a vital role in fiscal spending and taxation policymaking.

And inaccurate forecasts can lead to budget deficits, cash flow mismatches, or unplanned borrowing.

THE RESEARCH GAP:

Accurate forecasting of GST revenue remains a challenge due to the dynamic nature of economic activities, policy changes, tax evasion, and external macroeconomic factors such as inflation, interest rates, and consumer spending behavior.

Since the GST revenue model began in 2017, the limited availability of data makes the task more challenging.

APPLICATIONS

How an accurate revenue forcasting model can be applied:

Government Budget Planning: Improved revenue projections will help policymakers allocate funds more effectively for infrastructure, welfare, and development programs.

Anomaly Detection: By identifying deviations from expected trends, the model can serve as an early-warning system for economic shifts, compliance issues, or tax fraud.

Business and Policy Decision Support: Businesses can use forecasts to anticipate tax liabilities and plan cash flows, while policymakers can assess the impact of tax rate changes.

Literature Review

ECONOMETRICS

Forecasting Indian Goods and Services Tax revenue using TBATS, ETS, Neural Networks, and hybrid time series models

P.V. Thayyib, Muhammed Navas Thorakkattle, Faisal Usmani, Ali T Yahya & Najib H.S Farhan

Article: 2285649 | Received 27 Feb 2023, Accepted 15 Nov 2023, Published online: 03 Dec 2023

Cite this article https://doi.org/10.1080/23322039.2023.2285649 Check for updates

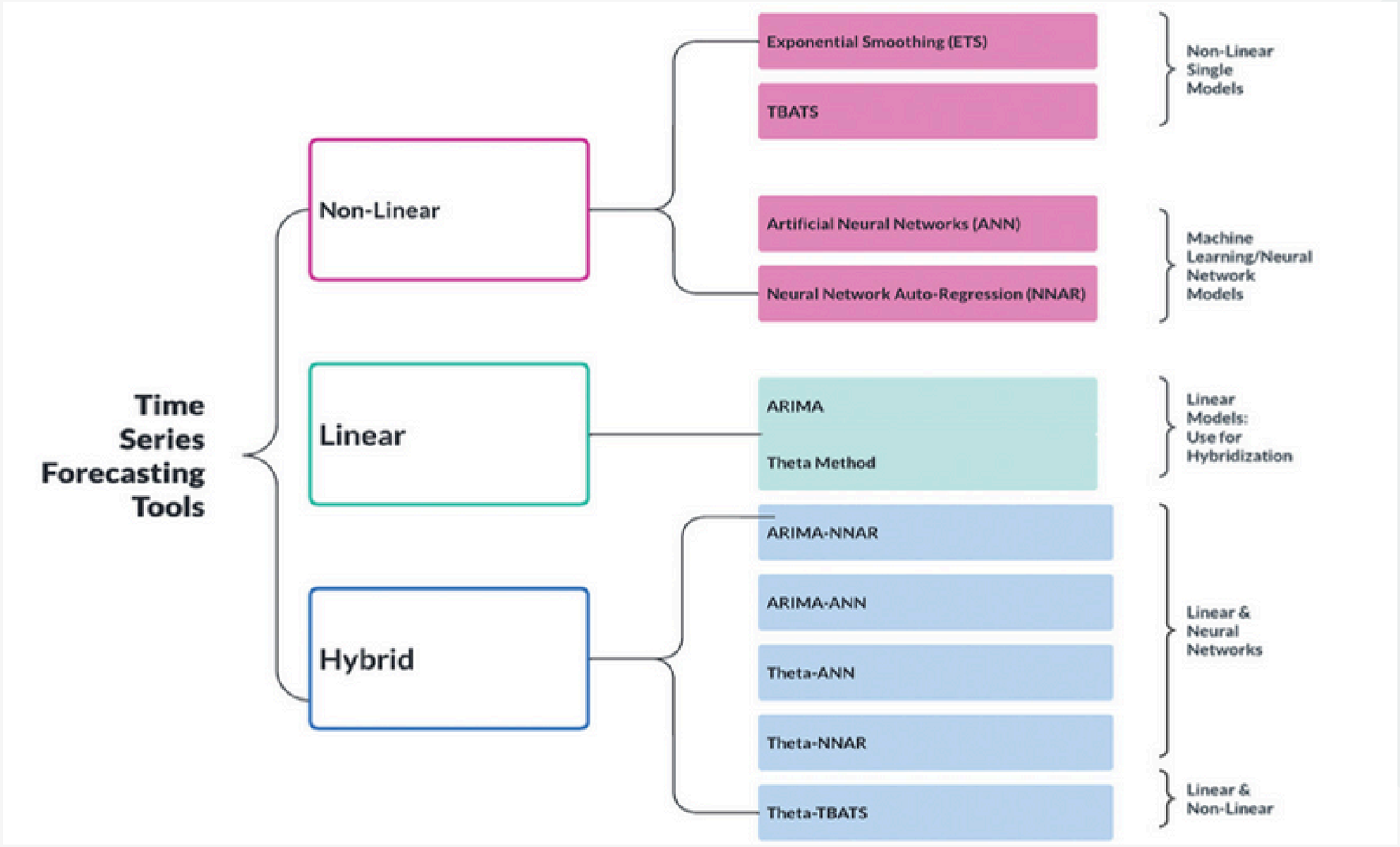
Context: The paper explores TBATS and ANN as a primary methods for revenue forcasting.

Dataset: Monthly unaudited Gross GST revenue [combining State GST, Central GST, Integrated GST (includes GST on Goods import), Compensation Cess (includes GST on Goods import)] revenue from the Indiastats Database, which was cross-verified at the GST Council Website. Considered monthly collection from August 2017 to November 2022

ML models: Explore alternative forecasting models, including Trigonometric Seasonality Box-Cox Transformation ARIMA errors Trend Seasonal components (TBATS) and Neural Networks: Artificial Neural Networks (ANN), Neural Networks for Autoregression (NNAR), which capture both linear and non-linear relationships

Limitation:

- Neural Networks May Not Be the Best for GST Forecasting
  - The study finds that Hybrid Theta-TBATS (a combination of a linear Theta model and non-linear TBATS) performs better than deep learning models like Neural Networks (NNAR).
  - Hybrid Theta-NNAR did not significantly improve accuracy, contradicting other studies (e.g., Bhattacharyya et al., 2022).
  - **Suggests that a "state-of-the-art" neural network may not always be robust for GST forecasting.**
- Lack of Exploration of Alternative Machine Learning Models
  - **The study does not test models like LSTM, CNNs, Random Forests, or Gradient Boosting, which could improve forecasting accuracy.**
  - **Future research should explore deep learning models, such as combining LSTM, XGBoost or using wavelet analysis and Bayesian workflow.**
- Limited Hybrid Model Combinations Tested
  - The study only tests Hybrid Theta-TBATS and Hybrid Theta-NNAR.
  - Other hybrid approaches (e.g., TBATS + ARIMA, Theta + LSTM) were not explored.
- Univariate Time Series Limitation
  - The models only use univariate GST revenue data, meaning they do not account for macroeconomic indicators, business cycles, or tax policy changes.
  - Future work should test multivariate models incorporating economic indicators (inflation, trade, consumption, etc.).


Time Series Forecasting Tools
Non-Linear
Linear
Hybrid
Exponential Smoothing (ETS)
TBATS
Artificial Neural Networks (ANN)
Neural Network Auto-Regression (NNAR)
ARIMA
Theta Method
ARIMA-NNAR
ARIMA-ANN
Theta-ANN
Theta-NNAR
Theta-TBATS
Non-Linear Single Models
Machine Learning/Neural Network Models
Linear Models: Use for Hybridization
Linear & Neural Networks
Linear & Non-Linear

Applying machine learning in tax revenue forecasting

Ching Hin (Jeffrey) Wong and Nathan La[1]

Department of Treasury and Finance

Context:
- Investigates ML vs. traditional models (e.g., AR(4), Random Walk) for payroll tax and land transfer dutyforecasting
- Uses 30 years of quarterly data and 9 forecasting algorithms
- Benchmarks performance under both normal and COVID-era volatility

Dataset:
- Payroll tax revenue
- Land transfer duty revenue
- Source: Victorian Department of Treasury and Finance (DTF).
- Time Period: June 1992 – December 2019 (expanded to September 2022 in some cases).

Features:
- Macroeconomic indicators (from the Australian Bureau of Statistics (ABS)).
- Property market indices (from CoreLogic).
- Total features: 23 in the baseline, expanded to 166 in some experiments.

“Machine learning methods are more effective for tax lines that have higher volatility and are more sensitive to economic fluctuations.”

“During abnormal periods like the COVID-19 crisis, machine learning models that explore nonlinear relationships — such as tree-based methods and neural networks — perform better.”

Limitations:

Machine Learning Wasn’t Always Effective:

- Traditional models like AR(4) (Autoregressive model) outperformed ML for payroll tax forecasting.
- ML models were only slightly better for land transfer duty (and only in volatile periods).

Limited Feature Selection Analysis:

- Uses 23 features initially, then 166 in an expanded test, but increasing features didn’t always improve performance.
- Adding too many features hurt model accuracy (suggesting feature selection techniques were needed).

No Real-Time or High-Frequency Data:

- Uses quarterly data, whereas GST revenues may require daily, weekly, or monthly predictions.
- Machine learning models may perform better with real-time data (social media, consumer spending, etc.), which wasn’t included.

Doesn’t Consider Policy Shocks:

- Tax revenues are heavily impacted by government policies (tax cuts, rate changes, exemptions, economic relief programs).
- No test was done to measure how policy announcements impact revenue forecasting.

A Comparison of LSTM, GRU, and XGBoost for forecasting Morocco's yield curve

Article *in* Mathematical Modeling and Computing · January 2024

DOI: 10.23939/mmc2024.03.674

Context: The paper focuses on time series forecasting using machine learning techniques to predict Morocco’s yield curve. It compares the performance of Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and eXtreme Gradient Boosting (XGBoost) models for forecasting the Moroccan Treasury bill reference rate. The study highlights the effectiveness of deep learning and machine learning techniques in financial time series prediction.

Dataset: The dataset used in the study consists of Moroccan Treasury bill reference rate data from July 1, 2015, to November 31, 2023, obtained from the Bank Al-Maghrib (Morocco's central bank).

ML models:

- LSTM (Long Short-Term Memory) – A type of recurrent neural network (RNN) designed to learn long-term dependencies in time series data.
- GRU (Gated Recurrent Unit) – A simplified version of LSTM with fewer parameters, aimed at improving efficiency and mitigating gradient vanishing issues.
- XGBoost (eXtreme Gradient Boosting) – A boosting-based decision tree model known for its efficiency and accuracy, especially in small datasets.

Limitations:

- Limited Economic Indicators:
  - Only uses Moroccan Treasury bill reference rate as input data.
  - Doesn’t incorporate macroeconomic indicators (GDP, inflation, trade data), which could improve accuracy.
- Time Series Assumptions:
  - Yield curve forecasting assumes stationarity, but economic factors often have long-term trends and shocks.
  - Deep learning models (LSTM, GRU) struggle with sudden policy changes or crises.
- Limited Exploration of Hybrid Models:
  - Uses LSTM, GRU, and XGBoost, but doesn’t test hybrid models like Transformer-based time series models (e.g., Temporal Fusion Transformers, Attention-based LSTMs).
- Small Dataset Issue:
  - 3,075 data points may be insufficient for deep learning models like LSTMs, which perform better on large datasets.
  - XGBoost outperformed deep learning likely due to the small dataset size.

“The XGBoost model outperformed all the other models in terms of forecasting performance.”

Table 5. MAE, MAPE, RMSE, and R^2 values of the three models.

ML Method	MAE	MAPE	RMSE	R^2
LSTM	0.001	0.043	0.00145	0.9571
GRU	0.00086	0.04	0.0011	0.972
XGBoost	0.00047	0.022	0.0007	0.9891

DATA!!

Features chosen as predictors, based on economic fundamentals and lit review:

- **Sensex30 Price:** Reflects economic activity, investor confidence, and corporate profitability, which influence GST collections.
- **RBI Interest Rates**: RBI interest rates affect borrowing, investment, and consumption, directly influencing GST revenue.
- **GDP:** A higher GDP indicates stronger economic activity, leading to higher consumption, production, and GST revenue.
- **Nifty_Pharma_Index**
- **FMCG_Pharma_Index**
- **Automobile_Pharma_Index**

**collected from govt websites and credible finance websites

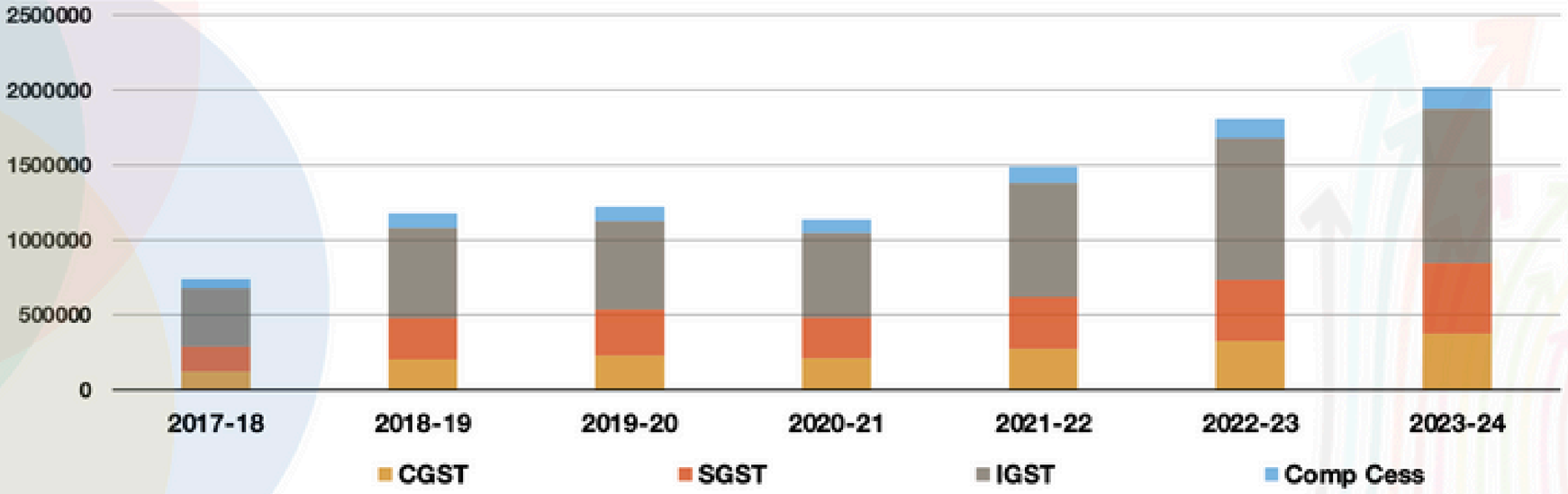
Month-Year	Timestamps	CGST	SGST	IGST	CESS	TOTAL	RBI_rate	Sensex	Price_of_USD_in_INR	Nifty_Pharma_Index	FMCG_Pharma_Index	Automobile_Pharma_Index
Jul-17	2017-07-01 00:00:00	10	10	20958	593	21572	6.25%	32,514.94	64.17	9476.40	25744.35	11002.65
Aug-17	2017-08-01 00:00:00	15252	23257	49968	7156	95633	6.00%	31,730.49	63.93	8859.65	25834.90	10612.55
Sep-17	2017-09-01 00:00:00	15131	21979	48930	8024	94064	6.00%	31,283.72	65.305	9172.60	24480.80	10811.25
Oct-17	2017-10-01 00:00:00	14962	22345	47995	8032	93333	6.00%	33,213.13	64.74	9756.00	25687.70	11370.00
Nov-17	2017-11-01 00:00:00	13690	20294	42694	7103	83780	6.00%	33,149.35	64.48	9238.75	25820.05	11292.90
Dec-17	2017-12-01 00:00:00	13927	19699	42765	7922	84313	6.00%	34,056.83	63.83	9620.10	26851.50	12009.70
Jan-18	2018-01-01 00:00:00	14870	21538	45338	8070	89817	6.00%	35,965.02	63.54	9384.75	27127.30	11611.90
Feb-18	2018-02-01 00:00:00	14757	20614	42381	8196	85947	6.00%	34,184.04	65.2	8960.25	26513.70	11157.20
Mar-18	2018-03-01 00:00:00	16257	22046	46326	7520	92148	6.00%	32,968.68	65.11	8358.05	26127.40	10821.35
Apr-18	2018-04-01 00:00:00	18647	25698	50548	8554	103448	6.00%	35,160.36	66.45	9060.70	28772.75	11625.75
May-18	2018-05-01 00:00:00	15862	21686	49119	7339	94005	6.00%	35,322.38	67.42	8220.25	28814.30	10994.45
Jun-18	2018-06-01 00:00:00	15968	22021	49498	8122	95610	6.00%	35,423.48	68.45	9173.15	28966.00	10709.05
Jul-18	2018-07-01 00:00:00	15877	22293	49951	8362	96483	6.00%	37,606.58	68.45	9179.40	31007.25	10973.75
Aug-18	2018-08-01 00:00:00	15303	21154	49875	7628	93960	6.00%	38,645.07	71.0	10390.95	32911.55	11009.25
Sep-18	2018-09-01 00:00:00	15318	21061	50070	7993	94442	6.00%	36,227.14	72.5	9972.45	29757.70	9590.25
Oct-18	2018-10-01 00:00:00	16464	22826	53419	8000	100710	6.00%	34,442.05	73.95	9757.50	28547.10	8820.55
Nov-18	2018-11-01 00:00:00	16811	23069	49725	8031	97636	6.00%	36,194.30	69.64	9275.70	30126.25	9270.20
Dec-18	2018-12-01 00:00:00	16442	22459	47936	7888	94726	6.00%	36,068.33	69.56	8868.70	30516.65	9235.55
Jan-19	2019-01-01 00:00:00	17763	24826	51225	8690	102503	6.00%	36,256.69	70.95	8825.35	29800.55	8218.40
Feb-19	2019-02-01 00:00:00	17625	24192	46953	8476	97247	6.00%	35,867.44	70.83	8884.85	29262.85	8355.15
Mar-19	2019-03-01 00:00:00	20353	27520	50418	8286	106577	6.00%	38,672.91	69.18	9346.55	30321.40	8335.35
Apr-19	2019-04-01 00:00:00	21163	28801	54733	9168	113866	6.00%	39,031.55	69.636	9402.50	30336.90	8350.60
May-19	2019-05-01 00:00:00	17811	24462	49891	8125	100289	6.00%	39,714.20	69.57	8455.10	29850.40	8175.50
Jun-19	2019-06-01 00:00:00	18366	25343	47772	8457	99939	6.00%	39,394.64	68.94	8065.15	29546.05	7928.05

We only have 90 data points, since its been only 7 years since GST’s implementation.

Payments- July'17 to Mar'24

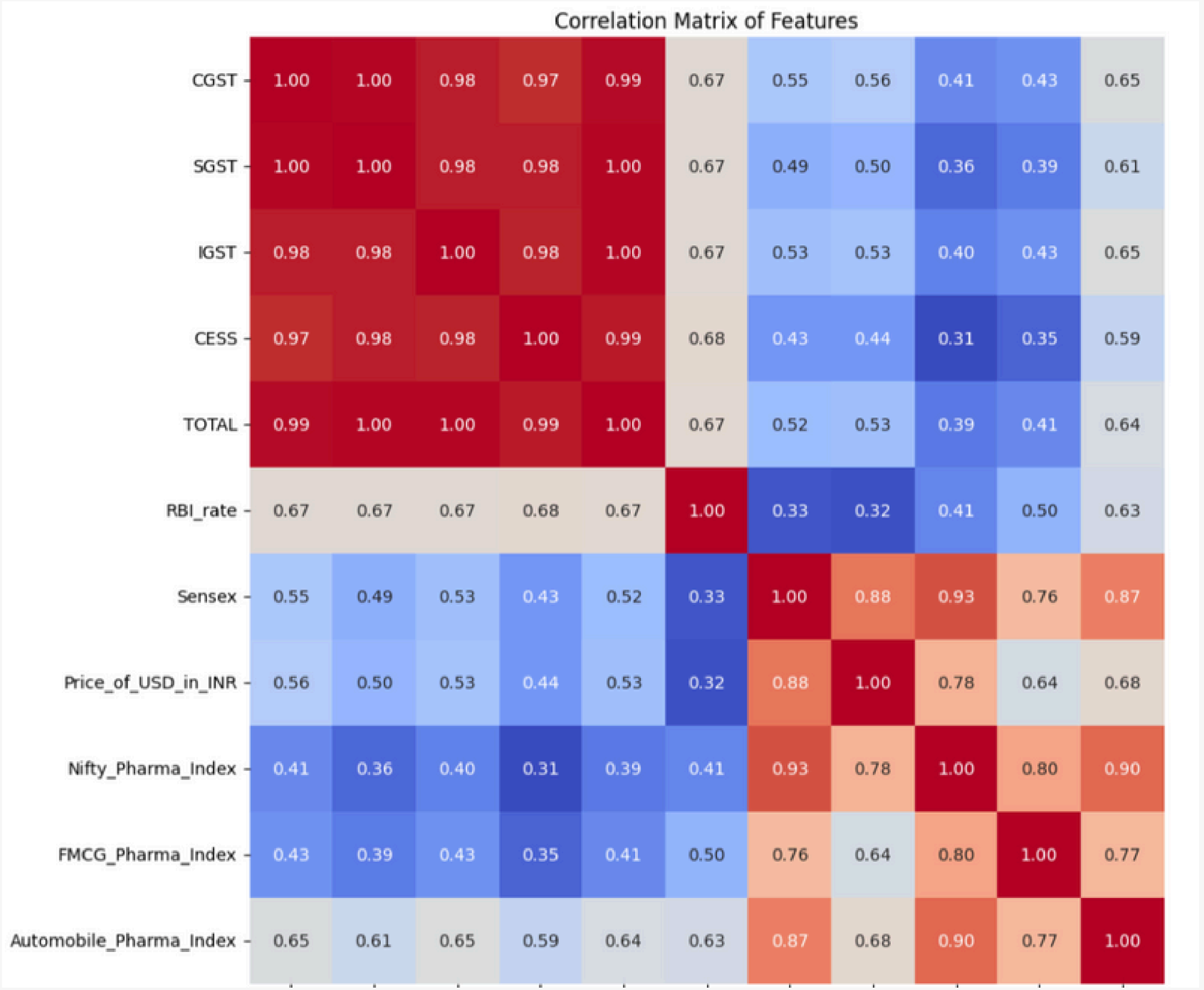
Figures in crores

MONTH	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23	2023-24
CGST	1,18,876	2,02,444	2,27,442	2,09,916	2,70,701	3,23,923	3,75,710
SGST	1,71,803	2,78,817	3,09,231	2,72,827	3,46,186	4,10,251	4,71,195
IGST *	3,87,355	5,98,739	5,86,698	5,65,720	7,63,632	9,45,220	10,26,789
Domestic	1,93,092	3,08,244	3,19,422	3,03,947	3,86,676	4,73,421	5,43,704
Imports	1,94,264	2,90,496	2,67,277	2,61,773	3,76,956	4,71,799	4,83,085
Comp Cess *	62,614	97,369	98,745	88,338	1,07,708	1,28,286	1,44,555
Domestic	56,319	87,289	88,304	79,153	98,918	1,17,390	1,32,639
Imports	6,294	10,079	10,443	9,185	8,790	10,896	11,915
Total	7,40,648	11,77,369	12,22,116	11,36,801	14,88,227	18,07,680	20,18,249



*Note -IGST/ Cess includes payments on both domestic supplies and imports

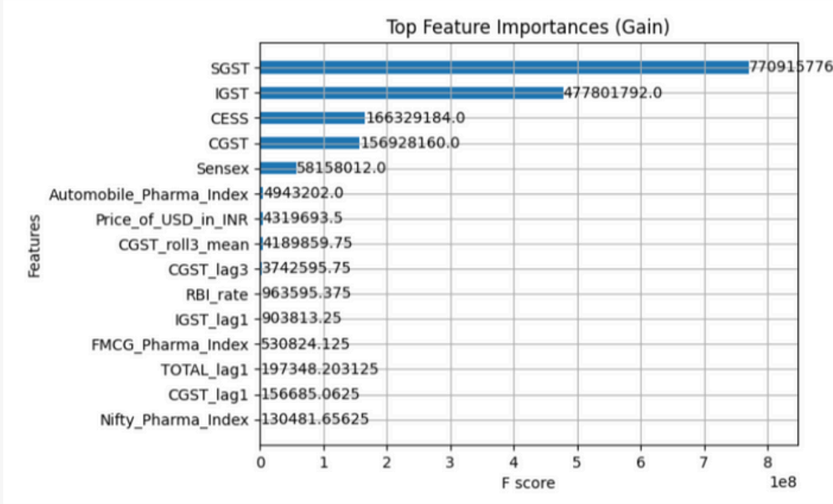
https://www.gst.gov.in/download/gststatistics


Correlation Matrix of Features
CGST 1.00 1.00 0.98 0.97 0.99 0.67 0.55 0.56 0.41 0.43 0.65
SGST 1.00 1.00 0.98 0.98 1.00 0.67 0.49 0.50 0.36 0.39 0.61
IGST 0.98 0.98 1.00 0.98 1.00 0.67 0.53 0.53 0.40 0.43 0.65
CESS 0.97 0.98 0.98 1.00 0.99 0.68 0.43 0.44 0.31 0.35 0.59
TOTAL 0.99 1.00 1.00 0.99 1.00 0.67 0.52 0.53 0.39 0.41 0.64
RBI_rate 0.67 0.67 0.67 0.68 0.67 1.00 0.33 0.32 0.41 0.50 0.63
Sensex 0.55 0.49 0.53 0.43 0.52 0.33 1.00 0.88 0.93 0.76 0.87
Price_of_USD_in_INR 0.56 0.50 0.53 0.44 0.53 0.32 0.88 1.00 0.78 0.64 0.68
Nifty_Pharma_Index 0.41 0.36 0.40 0.31 0.39 0.41 0.93 0.78 1.00 0.80 0.90
FMCG_Pharma_Index 0.43 0.39 0.43 0.35 0.41 0.50 0.76 0.64 0.80 1.00 0.77
Automobile_Pharma_Index 0.65 0.61 0.65 0.59 0.64 0.63 0.87 0.68 0.90 0.77 1.00

XGBOOST AND ITS FEATURES

Even though the correlation between GST revenues and non GST revenues is high (approximately 0.5), we initially assumed that they will have high feature importance in XGBoost as well.
However that wasn't the case because, XGBoost learns nonlinear, multivariate interactions, while correlation is linear and univariate.

- We experimented a lot with the features in order to decrease dimensionality, we tried to remove certain features and also applying PCA.
- However, it either lead to lower predictability or lower interpretability.
- Since our priority is predictability, we decided to use all the features. We might have to reduce dimensionality once our data set is big.
- A lag of 5 gave us the best results.

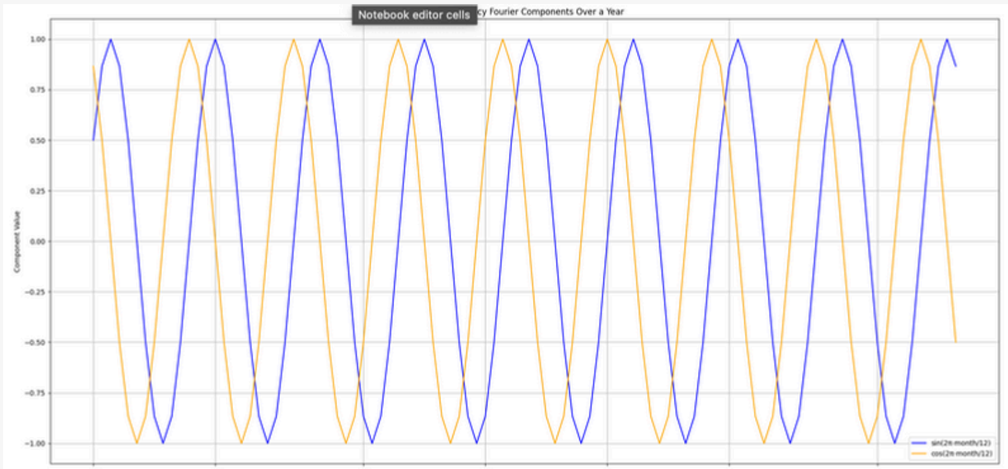


FOURIER TRANSFORMATION

Fourier transformation breaks down complex seasonal patterns into simple sinusoidal waves (sine and cosine curves). Instead of giving the model 12 separate month features, we capture repeating yearly or monthly behavior compactly.

- GST collections have strong monthly and yearly seasonality (e.g., higher collection near March, festive months, etc.).
- Traditional models like XGBoost don’t naturally understand time or seasonality.
- Fourier features inject this missing periodic signal, helping the model "sense" repeating patterns.
- Adding Fourier terms gives it explicit periodic signals → reduces unexplained variance → better fits seasonal swings→ higher R^2

The plot shows two low-frequency Fourier components (sin(2π·month/12) and cos(2π·month/12)) across one synthetic year.
These smooth, wave-like patterns represent repeating annual behavior.
The model learns to weigh these curves to reconstruct real seasonality in CGST trends.

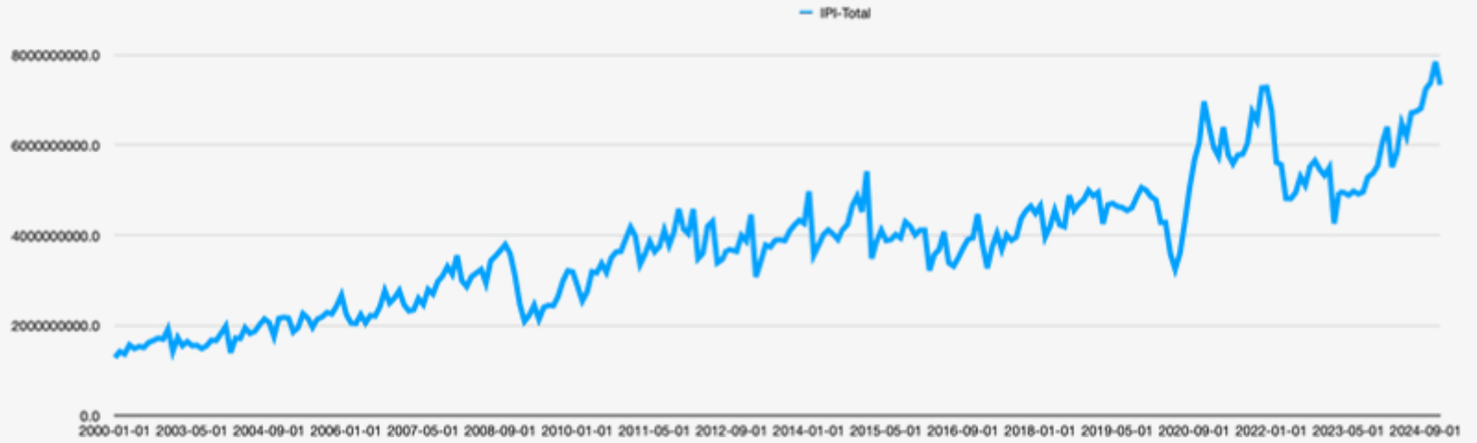


ADDING MORE DATA?

With only ~90 months of GST data in India, we used Brazil's similar tax structure as a proxy dataset. Testing both XGBoost and GRU on Indian and Brazilian data helped us compare models fairly, draw stronger inferences, and choose the best-fit approach for current and future forecasting.

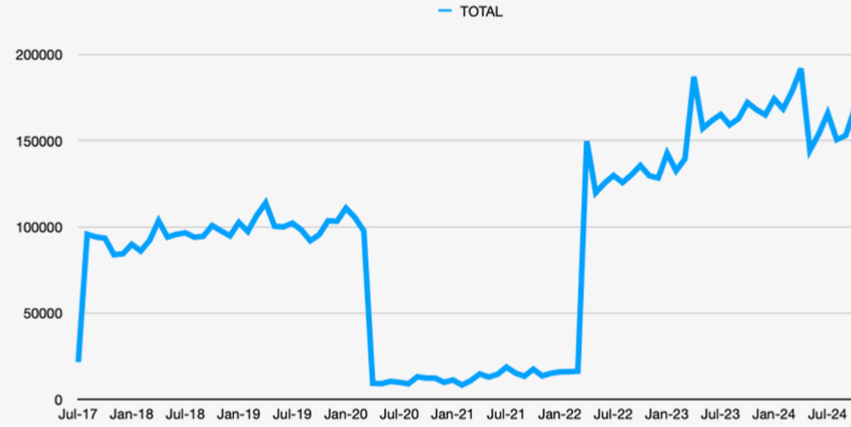
India implemented its dual GST model inspired by Canada and Brazil. However, due to the lack of accessible monthly revenue data from Canada, we selected Brazil.

Now let's **backtrack** and cover the lit review and dataset for Brazil....



The Imposto sobre Produtos Industrializados (IPI) is Brazil's federal tax on industrialized products. It is levied at the point of manufacture, import, or sale of goods and is broadly comparable to the indirect tax structure in GST.

The shortage of data and the dip due to covid in the total GST revenues clearly demonstrate a need for a proxy dataset



TAX STRUCTURE IN BRAZIL

1. Federal Level (Central Government)

Tax	Description
IPI	Tax on Industrialized Products (Excise tax); levied on manufactured goods
PIS	Program for Social Integration; payroll-based contribution
COFINS	Contribution for the Financing of Social Security; similar to a turnover tax

2. State Level

Tax	Description
ICMS	**State VAT on goods, electricity, transport, and communication**; varies by state

3. Municipal Level

Tax	Description
ISS	**Service Tax** collected by municipalities (Imposto sobre Serviços)

TAX STRUCTURE IN BRAZIL

Tax Type	Name	Jurisdiction	Characteristics
ICMS	Tax on Circulation of Goods and Services	State	Cascading tax, interstate disputes (fiscal war), credits difficult to use.
ISS	Tax on Services	Municipal	Varies by city, often overlaps with ICMS.
IPI	Tax on Industrialized Products	Federal	Applies to manufactured goods.
PIS / COFINS	Social Contributions on Revenues	Federal	Cumulative and non-cumulative versions, highly complex.

SIMILARITIES

- Federal Structure with Dual Taxation: Both countries operate under a federal system where taxation powers are divided between the central (federal) government and the states. In India, this is seen with the Central GST (CGST) and State GST (SGST); in Brazil, there are federal, state, and municipal taxes
- Value Added Tax (VAT) Principle: Each system is based on the VAT model, taxing value addition at each stage of the supply chain and allowing for input tax credits to avoid cascading taxes
- Multiple Tax Slabs: Both India and Brazil have multiple tax rates or slabs applied to different goods and services, rather than a single flat rate. Brazil has six main tax slabs (0%, 1.65%, 2%, 7%, 12%, and 17%), while India's GST features several slabs as well (0%, 5%, 12%, 18%, and 28%)

Afonso, J. R. R., Soares, J. M., & Castro, K. P. (2013). Evaluation of the structure and performance of the Brazilian tax system: White paper on taxation in Brazil (IDB-DP-265)

Doe, J., & Smith, A. (2025). GST Reform and Revenue Forecasting in Emerging Economies: Evidence from Brazil and India. SSRN.

SIMILARITIES

- Shared Tax Administration: In both systems, the central and state governments each have authority to levy and administer taxes on goods and services, reflecting a division of fiscal powers
- Destination-Based Taxation: Both India and Brazil have moved toward destination-based taxation, where tax is collected at the point of consumption rather than the point of origin, to reduce interstate tax competition and promote fairness in revenue distribution.
- Objective of Tax Unification: The overarching goal in both countries is to unify the indirect tax system, simplify compliance, and create a common national market by replacing a patchwork of previous taxes

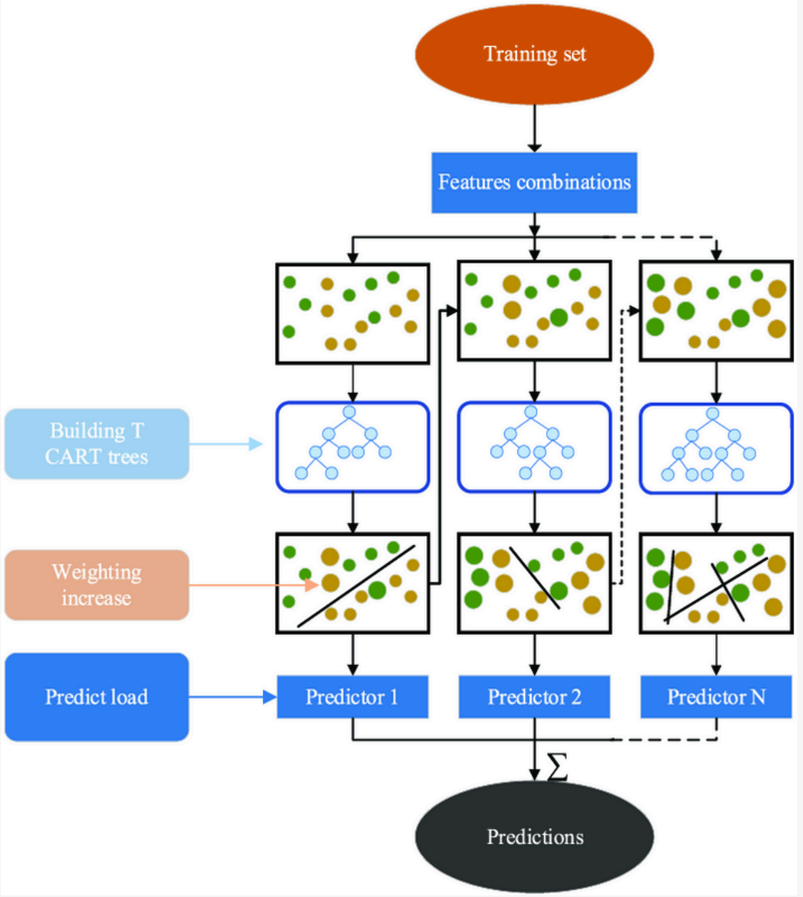
DIFFERENCES

Aspect	Brazil	India
Tax Structure Type	**Highly fragmented** multi-layered indirect tax system	**Dual GST model** – Central + State/UT taxes
Number of Taxes	Over **5 overlapping taxes**: IPI, ICMS, ISS, PIS, COFINS	Replaces over **17 indirect taxes** (excise, VAT, CST, service tax, etc.)
Jurisdictional Complexity	**Federal, state, and municipal** governments each levied and administered their own taxes separately	**Centre and states share tax administration** (CGST, SGST/UTGST, IGST)
Tax Base Differences	**Different tax bases and definitions** across taxes (e.g., goods vs. services)	**Uniform tax base** for goods and services across the country
Cascading Effect	No unified credit chain → **widespread tax cascading**	**Input Tax Credit (ITC)** across goods/services eliminates cascading
Place of Supply Conflicts	**Frequent tax wars** among states due to origin-based ICMS	GST is **destination-based**, reducing regional tax competition
Rate Uniformity	Different ICMS and ISS rates across states/municipalities	GST rates are **largely uniform**, decided by the **GST Council**
Revenue Distribution Mechanism	Complex, opaque sharing among levels of government	Clearly defined formula for **CGST-SGST revenue split**; IGST split by place of supply
Central Coordination Body	No unified decision-making forum; frequent legal disputes	**GST Council** ensures cooperative federalism and dispute resolution
Ease of Doing Business Impact	Complex system → **low EoDB rankings**, discouraged formalization	GST implementation improved India's **EoDB ranking significantly**

HOW XGBOOST WORKS

EXtreme Gradient Boosting, is an advanced machine learning algorithm designed for efficiency, speed, and high performance. It extends traditional gradient boosting by including regularisation elements in the objective function, XGBoost improves generalisation and prevents overfitting.

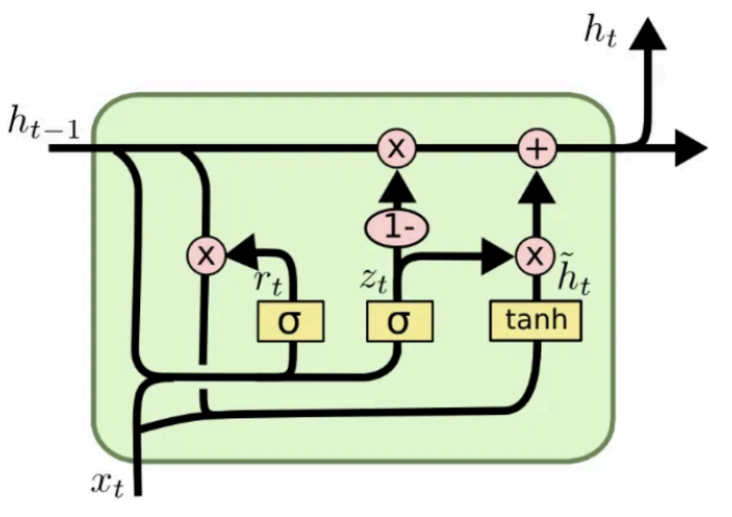
1. Start with a base learner: The first model decision tree is trained on the data. In regression tasks this base model simply predict the average of the target variable.
2. Calculate the errors: After training the first tree the errors between the predicted and actual values are calculated.
3. Train the next tree: The next tree is trained on the errors of the previous tree. This step attempts to correct the errors made by the first tree.
4. Repeat the process: This process continues with each new tree trying to correct the errors of the previous trees until a stopping criterion is met.
5. Combine the predictions: The final prediction is the sum of the predictions from all the trees.



HOW GRU WORKS

GRU (Gated Recurrent Unit) is a neural network component designed to remember important information over long sequences.
GRUs aim to simplify the LSTM architecture by merging some of its components and focusing on just two main gates: the **update gate** and the **reset gate**, while also being better for smaller datasets

1. Start with input and previous memory: GRU receives both new information and what it remembered from before.
2. Update Gate decides what to keep: This gate determines how much of the previous memory should be retained (like remembering a key detail from earlier in a conversation).
3. Reset Gate filters old information: This gate decides which parts of the previous memory are no longer relevant (like forgetting unimportant details).
4. Create new memory candidate: GRU combines the filtered old memory with new information to create a potential new memory.
5. Final memory update: The update gate blends the previous memory and the new candidate to form the final memory that's passed forward.



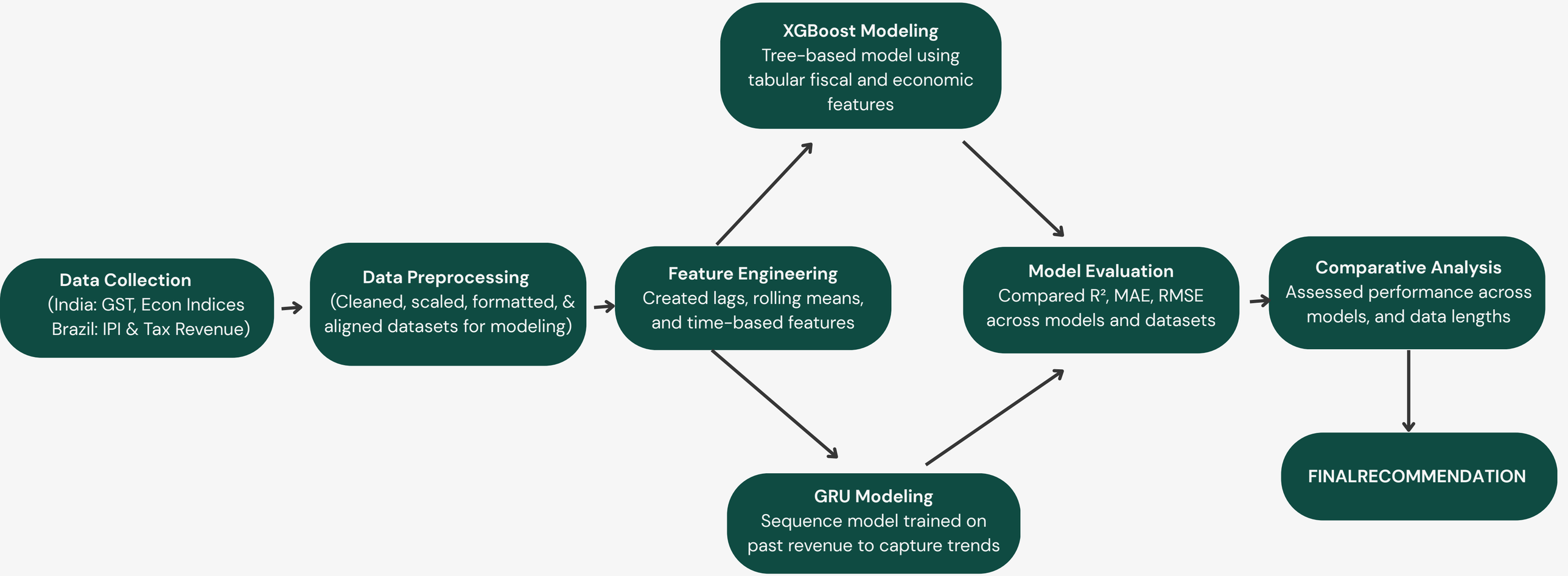
$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

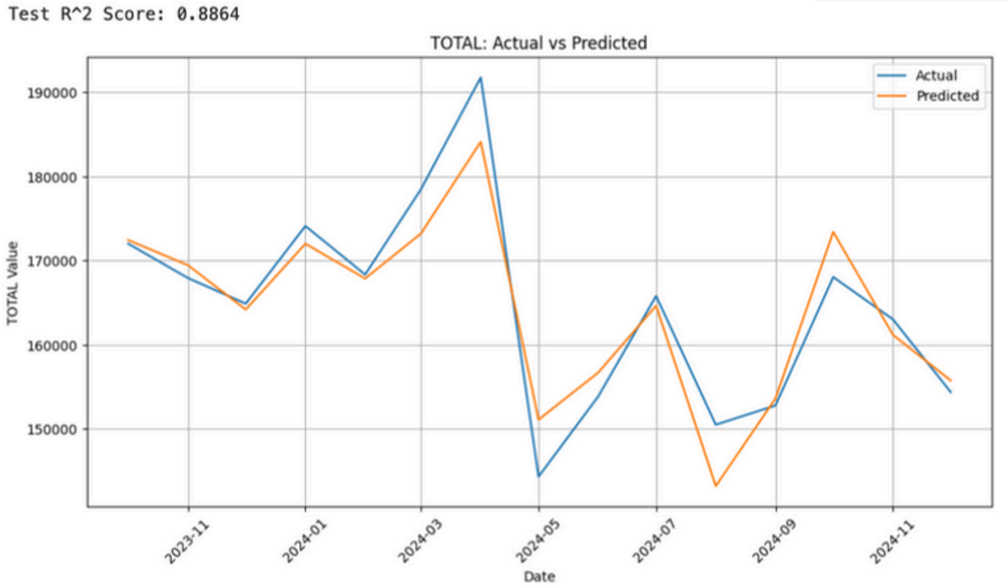
ML Methodology



PERFORMANCE METRICS

After applying XGBoost and GRU on both datasets we achieved these values..
We will only be comparing R^2 values, as MSE and RMSE wont be comparable due to difference in currencies and units.

Model	R^ 2
XGBoost for Total GST and CGST	0.8864, 0.8113
GRU for Total GST	0.317
XGBoost for IPI (Brazil)	0.943
GRU for IPI (Brazil)	0.723



GRU's performance significantly improves as we move to 300 data points. However it still doesnt outperform XGBoost

DEPLOYABILITY RECOMMENDATIONS

If this system were to be deployed in India today, we recommend using XGBoost as the forecasting engine.

- It is efficient, easy to update, and interpretable — ideal for integration into government dashboards or fiscal planning tools.
- As India's GST database grows over the next 20–25 years, deep learning models like GRU may become more viable.
- At that point, a transition to hybrid or GRU-based systems can be considered to better capture complex trends.

CHALLENGES

If this system were to be deployed in India today, we recommend using XGBoost as the forecasting engine.

- It is efficient, easy to update, and interpretable — ideal for integration into government dashboards or fiscal planning tools.
- As India's GST database grows over the next 20–25 years, deep learning models like GRU may become more viable.
- At that point, a transition to hybrid or GRU-based systems can be considered to better capture complex trends.

DATA COLLECTION

- GST Council, Government of India.
  https://www.gstcouncil.gov.in/
- Reserve Bank of India (RBI).
  https://www.rbi.org.in/
- Ministry of Finance, Government of India.
  https://finmin.nic.in/
- Brazilian Federal Revenue Service (Receita Federal).
  https://www.gov.br/receitafederal/pt-br

BIBLIOGRAPHY

- GST in India - https://doi.org/10.2991/978-94-6463-696-3_23

- Thayyib, P. V., Thorakkattle, M. N., Usmani, F., Yahya, A. T., & Farhan, N. H. S. (2023). Forecasting Indian Goods and Services Tax revenue using TBATS, ETS, Neural Networks, and hybrid time series models. Cogent Economics & Finance, 11(2). https://doi.org/10.1080/23322039.2023.2285649

- Doe, J., & Smith, A. (2025). GST Reform and Revenue Forecasting in Emerging Economies: Evidence from Brazil and India. SSRN. https://ssrn.com/abstract=5079698

- IJLMH. "GST in India: Lessons from Canada and Brazil."
- Jha, Nimisha. "GST in India: Lessons from Canada and Brazil." International Journal of Law Management & Humanities, vol. 8, no. 1, 2025, pp. 1315-1321. ISSN 2581-5369. DOI: https://doij.org/10.10000/IJLMH.119011.

- Mahajan, G. (2025). One Nation, One Tax: Evaluating Revenue Performance of the Indian GST System. SSRN. https://ssrn.com/abstract=5079698

BIBLIOGRAPHY

- Wong, C. H. (Jeffrey), & La, N. (2024). Applying machine learning in tax revenue forecasting. Victoria's Economic Bulletin, 8(2), May 2024. Department of Treasury and Finance, Victoria. https://www.dtf.vic.gov.au/sites/default/files/2024-09/Victoria%E2%80%99s-Economic-Bulletin,-Volume-8,-Number-2,-May-2024.pdf

- Sharma, N. S., & Dayama, V. (2023). A study on comparative analysis of GST revenue. B. V. Patel Institute of Commerce, Uka Tarsadiya University, Bardoli, India.

- Dandona, I., Tomar, P. K., Gupta, S. K., & Verma, S. K. (2024). GST dynamics in India: Exploring state revenue trends, GDP impact, and economic resilience. Malque Research, [Issue/Volume if available], Article 2889. https://malque.pub/ojs/index.php/mr/article/view/2889

- Degife, W.A.; Lin, B.-S. Deep-Learning-Powered GRU Model for Flight Ticket Fare Forecasting. Appl. Sci. 2023, 13, 6032. https://doi.org/10.3390/app13106032

- Author(s). (2024). A Comparison of LSTM, GRU, and XGBoost for forecasting Morocco's yield curve. Mathematical Modeling and Computing, January 2024.

THANK YOU