Machine Learning Methods For Bilateral Trade Flow Prediction Between Global Economies

Prateek Rao
PES1UG20CS303, Dept. of CSE
PES University
Bangalore, India
xrprateek@gmail.com

Rahul Ramesh
PES1UG20CS319, Dept. of CSE
PES University
Banglore, Karnataka
ynotrr@gmail.com

Pooshpal Baheti
PES1UG20CS283, Dept. of CSE
PES University
Banglore, Karnataka
pooshpalbaheti@gmail.com

***Abstract*—Bilateral trade flow between global economies has been a key economic indicator for economists and policymakers as they can significantly influence international trade sanctions and policies, which will undeniably affect international trade relationships. This global interdependence has been historically stimulated by the international economy which establishes substantial mutual benefits for the participating countries. The concept of analysing various economic trends and making accurate future predictions has become indispensable since the advent of machine learning models. By scrutinising historic economic data and building relevant ML models, we can attempt to predict future events and capital flow between participating countries. It goes without saying that the accurate prediction of bilateral flow is of paramount importance and the use of machine-learning approaches for economic forecasting, over the traditional statistical methods, can help in achieving exceptional predictions. In this paper, we attempt to improve upon traditional statistical methods, by experimenting with different machine learning models.**

I. Introduction

Policymakers for international trade and global trade representatives often work hand in hand with economists to delve into the various economic indicators to formulate strategies, policies, reforms, and roadmaps for crucial economic decisions. Bilateral trade flow has historically been a substantial indicator of global interdependence between various countries. It alludes to the international flow between two countries. Broadly put, it represents the value of goods and services that have been exported from one country to another, influencing international trade policies in the participating countries.

Economists can make numerous inferences by using bilateral flow between economies as an indicator of international economic relations. These inferences represent economic stability, development, and growth which are often substantial for policymakers to derive adequate agendas to meet fiscal goals.

A higher rate of flow between countries indicates a developing relationship between the participating economies. Taking the viewpoint of developing countries, an increase in imports will create opportunities for employment, as well as allows for greater resources, while also expanding the competitive markers. An increase in exports will significantly increase the GDP of a country which will reduce the existing poverty gap. Bilateral trade flow prediction can help policy makers of the G20 countries to take more informed decisions and alter existing strategies to better them.

II. Related Work

We first focus on the revolutionary work of Walter Isard in 1954, [1] which gave rise to the gravity model of international trade. This model predicts bilateral trade flows based on the economic sizes and distance between two units. Between two countries i and j, it takes the form of

$$F_{ij} = G.\frac{M_i^{\beta_1} M_j^{\beta_2}}{D_{ij}^{\beta_3}}$$

Here, G is a constant, F stands for trade flow, D stands for distance and M stands for economic dimensions of the countries that are being measured. The gravity model has been used throughout history to predict some important events, and it provides us with a suitable baseline for evaluation. Independend calculations have suggested that the gravity model has an adjusted R^2 value between 0.5 and 0.6 .

Isaac Wohl and Jim Kennedy [2] used artificial neural networks to analyze international trade data. Making use of data between United States and it's main trade partners, they trained the model on the set from 1986 to 2006, and made predictions from 2007 to 2016. Results were reasonably close to the real values.

Jingwen Sun et al. [3] analyzed Bilateral Trade Flow, but with a different target variable in mind. They used yearly import/export data from 217 countries for the 1960-2017 period, and predicted the GDP using several statistical factors. They concluded that RBF Regressor was the better model.

S Circlaeys et al. [4] tested several machine learning models to predict pattern of bilateral trade flows. They used a variety of linear regression and neural network models to train on economic and geographic factors. Due to the ability of neural networks in capturing non-linear interaction among features, it performed better than any regression-based model.

Feras Batarseh et al. [5] used machine learning algorithms to select the best economic variables that influence trade between specific commodities, and train models using these features.

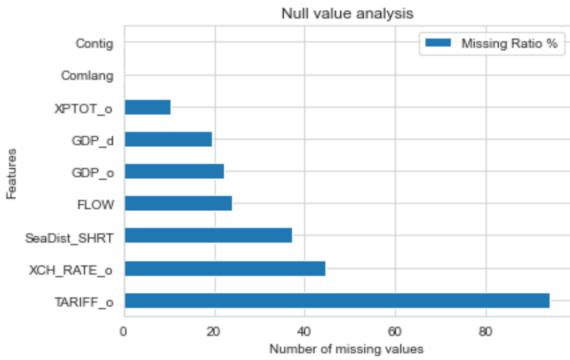

Fig. 1. Null Value Analysis

This type of feature engineering was found to be more accurate than statistical methods in predicting future trend patterns. Jin-kyu Jung et al [6] employed a number of ML methods as a new approach to forecast. They applied the Elastic Net, SuperLearner, and Recurring Neural Network on macro data of seven advanced and emerging economies. Results show that these ML algorithms outperformed conventional statistical approaches and offering great potential in the area of economic forecasting.

III. Dataset

We intend to make use of the "TradHist" dataset by CEPII [7], which provides around 1.9 million bilateral trade observations for 1827-2014, for 200+ different countries and regions. It has both geographical dimension and time dimension. Due to the huge time-frame of this data, there are a lot of missing values due to difference in data collection techniques. We employ a lot of data cleaning and preprocessing steps in order to prepare the data accordingly for our use cases.

A. Data Cleaning

In Fig. 1., we can see the null value ratio in most of the important features. These null values can cause a lot of issues during data analysis and training. Hence, we start off by removing data with any null values in our feature columns. We also remove any numeric column with zero values, as this can potentially introduce unnecessary noise into our training data. Detection of outliers was conducted and treatment was done on the dataset likewise. We also focus on recent data only, because we intend to forecast recent trends. We also get rid of any trade flow value which is lesser than 100 Pound Sterling, as it is very insignificant.

B. Exploratory Data Analysis

In this section, we will explore the data briefly in order to understand trends and uncover some patterns. First, let us examine the number of bilateral trade observations for different values of `FLOW`, which serves as our target variable.

In Fig. 2., we immediately notice that the number of observations is increasing per year, which can be attributed to three factors: First, it reflects the increasing number of existing countries, which mechanically increases the number of potential international trade flows. Second, the increasing number of observed flows reflects the increasing number of existing country pairs that are actually engaged in bilateral trade. Third, it can be a consequence of the easier access to primary sources in recent years, due to conservation issues and the difficulties in locating historical statistics for more ancient times.

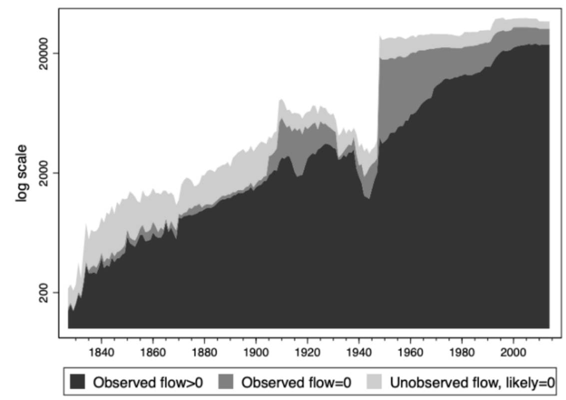

Fig. 2. Number of bilateral trade observations

On performing Correlation Analysis, we notice that there are some strong linear correlations between our data, as outlined in Fig. 3. We can observe that there is a significant linear positive correlation between the GDP of the origin country, the GDP of the destination country, and the total exports of the origin country, with respect to the target variable `FLOW`. We will take these relationships into consideration as we move to feature selection.

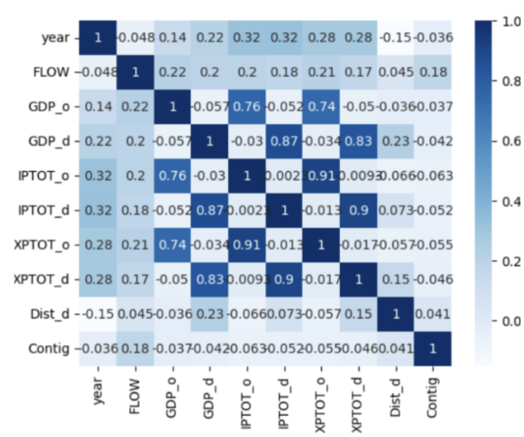

Fig. 3. Correlation Heatmap

C. Feature Engineering

In order to further explore relationships between data, we run a preliminary round of LightGBM on our prepared data, and plot the feature importances in predictions, as outlined in Fig. 4.

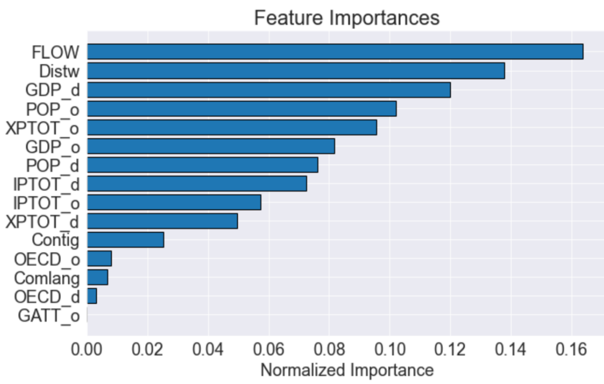

Fig. 4. Normalized feature importances

On further inspection, we don't find an issue of spurious correlation here. Thus, we will incorporate the top 12 features into our further working models, and use these as exogenous features to model our predictions.

Although we realize that we have passed up an opportunity to perform advanced feature engineering by construction of our own features, the aim of this paper is to compare the performance of machine learning models to the Gravity Model, and hence we will operate on a similar feature space as the Gravity Model.

D. Data Transformation

On plotting the histogram for the target variable, we find that the data is highly skewed. In order to bring it to a normal distribution, we use a log transformation on all the numeric attributes of the dataset. In Fig. 5., we can observe that the log transformation is useful in normalizing the distribution of the target variable.

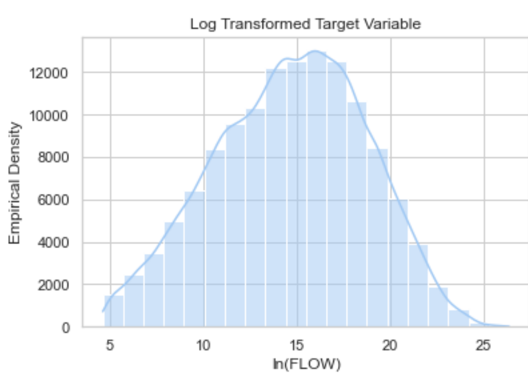

Fig. 5. Log Transformed Distribution

E. Time Series Transformation

All of the above data transformation methods are used to prepare the data for time-agnostic modelling. However, we also explore time series modeling for this data. Here, we take a pair of countries as our tuple, which acts as the unique identifier. We cut the data down to between 2001 and 2014 only, as we want to explore recent trends. Log transformations are applied to normalize the data.

IV. Proposed Methods

In order to select the best model, we have tried and tested multiple models. These models have been highlighted in the following section. Although most of the models make use of time agnostic features, we have attempted to transform the given data into time series data and build the models with time lagged features. The premise of this paper is to highlight how time agnostic models can generalise and predict a target that would generally be considered as time series.

We provide some insight into our model selection process in this section. Since we are predicting `FLOW`, which happens to be a continuous variable, we will focus on implementing regression-based models over classification. On data exploration, it is obvious that most of the interactions between features are non-linear, and hence we will be focusing on models that can best capture non-linear correlations. We have been careful to avoid overfitting, and each model has been tested against completely unseen data. Before training of the model, the data has been split into the usual Train-test-validation. We decided not to use K-fold Cross Validation as we have ample number of data instances to test against.

The first model to be tried was normal Linear Regression, which accounts for the best fit line. The model regresses the bilateral trade flow data as such:

$$FLOW = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n + \epsilon$$

However, as expected, the model performed horribly as it wasn't able to capture non-linear interactions. We proceeded to try out Log Transformed Linear Regression, with the entire numerical dataset being transformed by Logarithmic function to distribute the data normally.

$$FLOW = \beta_1 ln(GDP_o) + \beta_2 ln(GDP_d) + ... + \beta_n ln(Distw) + \epsilon$$

This model provided a slightly better result, with a R^2 value of 0.65, which provides a good baseline to compare.

The third model tried was a Support Vector Machine, with logarithmic features. Here, we employ the RBF kernel, to try to model non-linear features.

$$k(x, x') = \exp(\frac{-(x - x')^2}{2\gamma^2})$$

We also implement a fully feed-forward artificial neural network, without hyper parameter tuning. The model architecture is shown in Fig. 6. We make use of logarithmic features and aptly scaled values to train the neural network. We make use of both tanh and relu activation functions.

The last time agnostic model implemented was LightGBM, which is an ensemble gradient boosted model. It makes use of decision trees to regress the predictive variable. The number of estimators chosen was 10000.

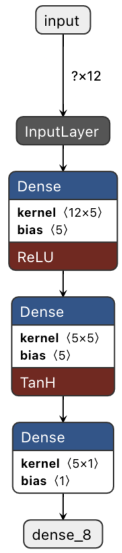

Fig. 6. ANN Model architecture

We now move toward time series modelling, by making use of the data transformed specifically for time series. We can only utilize the data between two countries.

The first time series model implemented was ARIMA(2,1,2). The order was calculated by plotting the ACF and PACF plots.

We further implemented ARIMAX(2,1,2) leveraging exogenous variables. We implemented Augmented Dickey-fuller test and discovered that we can obtain a stationary series through first order differenciation.

V. Results

A. Comparison Metric

To compare the predictive performances of the models, we used the coefficient of determination (R^2) as the comparison metric. It can be formulated as shown below:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

B. Validation Technique

To eliminate any sort of bias, all the models have been validated in the same format. We have a 30% holdout of the total data for testing and validation. As mentioned earlier, k-fold cross validation was not considered as we had enough data. caption

Table 1: Model Performances

Models	R^2
Linear Regression	0.14
ARIMA(2,1,2)	0.38
ARIMAX(2,1,2)	0.61
Log Transformed Linear Regression	0.65
Support Vector Regression	0.67
Artificial Neural Network	0.70
LightGBM	0.91

C. Inferences

Here, we can observe that the LightGBM Model is performing the best, with an R^2 value of 0.91, which is state-of-the-art. We infer this is because Gradient Boosting models are exceptional with tabular data, and making use of the feature analysis, we are able to capture most of the non-linear interactions between the independent variables.

Although the Artificial Neural Network model seems to be underperforming, there is a lot of scope for hyperparameter tuning, to make the model perform better.

Fig. 8. shows the predicted vs. actual values for Flow, with USA as source country, using the LightGBM model. As observed, the predictions are pretty spot-on in most cases. Special care was taken to make sure that there was no data leakage.

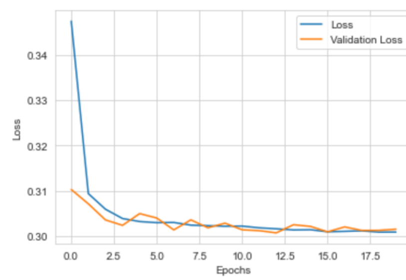

Fig. 7. ANN Training

VI. Conclusion

Our results show that using a stochastic gradient boosting model such as LightGBM shows promising results in modeling tabular data and capturing non-linear relationships. In theory, neural networks should also provide similar results due to it's advancements in modeling non-linear relationships, and it is an option worth being explored further.

Time-series models provided disappointing results due to the fact that there did not exist enough data to accurately model between two countries. Utilizing deep neural networks for time series data would've resulted in overfitting, and hence was skipped.

VII. Future Work

As denoted by our results, gradient boosting with hyperparameter tuning is worth exploring further. Tuning ANNs is also another avenue. More advanced neural network models such as CNNs can also be explored. Continuing with time series modeling would be useful if more data per country could be collected. This could open up avenues to explore RNNs and LSTM models as well.

Acknowledgment

We would like to acknowledge our Data Analytics Course Professor Mr. Anand MS for providing constant guidance during each phase of our project. We would also like to acknowledge the course anchor Dr. Gowri Srinivasa, who has prepared the course content and also the teaching assistants

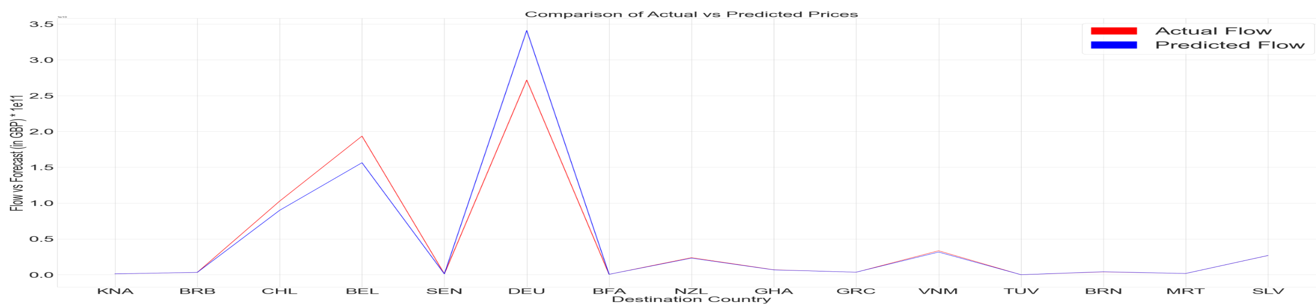

Fig. 8. Predicted Flow with Source Country USA for 2013

who have been constantly providing resources to practice the learnt concepts.

Contribution of Team Members

- Prateek Rao - Data Preprocessing, Machine Learning Models, Inferences
- Rahul Ramesh - Data Preprocessing, Analysis.
- Pooshpal Baheti - Exploratory Data Analysis, Inferences

Peer-review Addressal

We have implemented and tested several more models to increase accuracy, as required by peer review.

References

[1] W. Isard, "Location theory and trade theory: Short-run analysis," *The Quarterly Journal of Economics*, vol. 68, no. 2, pp. 305–320, 1954.

[2] I. Wohl and J. Kennedy, "Neural network analysis of international trade," *US International Trade Commission: Washington, DC, USA*, 2018.

[3] J. Sun, Y. Suo, S. Park, T. Xu, Y. Liu, and W. Wang, "Analysis of bilateral trade flow and machine learning algorithms for gdp forecasting," *Engineering. Technology & Applied Science Research*, vol. 8, no. 5, pp. 3432–3438, 2018.

[4] S. Circlaeys, C. Kanitkar, and D. Kumazawa, "Bilateral trade flow prediction," *Unpublished manuscript, available for download at http://cs229. stanford. edu/proj2017/final-reports/5240224. pdf*, 2017.

[5] F. Batarseh, M. Gopinath, G. Nalluru, and J. Beckman, "Application of machine learning in forecasting international trade trends," *arXiv preprint arXiv:1910.03112*, 2019.

[6] J.-K. Jung, M. Patnam, and A. Ter-Martirosyan, *An algorithmic crystal ball: Forecasts-based on machine learning*. International Monetary Fund, 2018.

[7] M. Fouquin and J. Hugot, "Two centuries of bilateral trade and gravity data: 1827-2014," Working Papers 2016-14, CEPII, May 2016.