ISYE 6740 – Computational Data Analysis - Summer 2023
Final Report

Team Member: Sammy Rabby

Project Title: Sovereign Credit Rating Forecasting using Random Forest Classifier and Deep Neural Networks

1. Problem Statement

Sovereign credit rating is an assessment of a sovereign's creditworthiness considering both its capacity and willingness to pay its credit obligations. Credit ratings in general give investors and credit lenders an idea of the level of risk associated with investing or lending to the corresponding entity. It is also a critical measure in financial risk management as the level of capital held by a financial institution for a given exposure is dependent on the exposure's credit rating.

The objective of this project is to analyze the performance of Random Forest Classifier and Deep Neural Networks such as MLP, RNN, LSTM, and CNN on forecasting the next quarter sovereign credit ratings with Standard & Poor's (S&P) as the benchmark. For decades, credit assessments of the top rating agencies such as S&P, Moody's, and Fitch have not changed fundamentally. Rating agencies use a combination of qualitative and quantitative analysis. There are subcategories under the qualitative and quantitative analysis. It then uses a weighted approach in order to arrive at a single measure translated to the sovereign's credit rating. However, for this work, I will only consider quantitative economic variables such as GDP, Inflation, Government Debt as % of GDP, etc. For future improvements, qualitative or categorical variables may be considered.

There is literature on using Neural Networks[1] and machine learning techniques such as multilayer perceptron (MLP), classification and regression trees (CART), support vector machines (SVM) to model sovereign credit ratings[2] however I haven't seen usage of RNNs and LSTM yet. There is usage of RNNs and LSTM for credit rating assessments[3] however it was only used for corporate credit ratings which is entirely different from sovereign credit ratings.

2. Data Source

The main data sources for this project will be IMF's International Financial Statistics (IFS)[1] and World Bank's Data Bank [2]. Both databases contain over 1000 economic

[1] León-Soriano, R., Muñoz-Torres, M.J. (2012). Using Neural Networks to Model Sovereign Credit Ratings: Application to the European Union

[2] Overes, B.H.L., van der Wel, M. Modelling Sovereign Credit Ratings: Evaluating the Accuracy and Driving Factors using Machine Learning Techniques

[3] Florescu, I., Golbayani, P., Wang D. Application of Deep Neural Networks to assess corporate Credit Rating.

indicators. Both of which are publicly available. IFS Data can be extracted via Query Tables or JSON RESTful Web Service while World Bank Data can be exported to excel. For now, I have extracted data into excel .xlsx via Query Tables.

For this study, I chose 24 economic variables from different categories. This ensures that different economic aspects are represented in the model. Table 1 below shows all 24 economic variables chosen for this study and its corresponding category.

There could still be other relevant variables that were not included such as Reserves held and Debt Service. The reason is that this data is missing for a lot of countries in the main data sources or is not in the form that can be quickly used for the study such as being in local currency. For further improvements, data wrangling can be performed on these highly important economic variables to test if there is improvement in the model.

Table 1: Economic Variables used as Features.

Category	Economic Variable
Economic Concentration	Agriculture, forestry, and fishing, value added (% of GDP)
	Manufacturing, value added (% of GDP)
	Services, value added (% of GDP)
	Trade (% of GDP)
	Trade in services (% of GDP)
Debt Level and Investment Flows	Central government debt, total (% of GDP)
	Current account balance (% of GDP)
	External balance on goods and services (% of GDP)
	Foreign direct investment, net inflows (% of GDP)
	Foreign direct investment, net outflows (% of GDP)
Inflation and Market Factors	Inflation CPI
	TBills - Yield
GDP and Components	GDP per capita (constant 2015 US$)
	Expense (% of GDP)
	Gross savings (% of GDP)
	Exports of goods and services (% of GDP)
	Imports of goods and services (% of GDP)
	General government final consumption expenditure (% of GDP)
Governance Indicators	Control of Corruption: Estimate
	Government Effectiveness: Estimate
	Political Stability and Absence of Violence/Terrorism: Estimate
	Regulatory Quality: Estimate
	Rule of Law: Estimate
	Voice and Accountability: Estimate

An important point on choosing the features is that they should be comparable across countries and periods (stationary for time-series). Hence, they are usually normalized as percentage of GDP or per Capita. A lot of this issue are can be seen in the IMF data. Despite its advantage of being quarterly, most of its data are in local currency and not normalized, hence the difficulty to process it. For future improvements, maybe some data transformations can be done to it as I think it will really be useful inputs to the model.

For the labels, I will be using Sovereign Credit Ratings produced by S&P. I will specifically use the sovereign's Long-Term Foreign Currency Rating. Two sets of models will be trained every time to correspond to a label with No Rating Modifiers and one with Rating Modifiers (+ and – to letters).

Other supplementary data used is the ISO 3166-1 Alpha-3 Codes of countries. This will be used to standardize the naming conventions between the IMF, World Bank, and S&P Data.

3. Data Wrangling and Preparation

The 24 selected features are the result of several iterations of data wrangling and preparation in order to arrive to reasonable amount of data points. Having more features will result to a lot of missing data as it is not always available for other countries. I didn't do any data imputation for this project. For future improvements, that may be a focus point as including more relevant economic variables that is missing for some countries may help the model.

The goal is to combine World Bank and IMF features into one Data Frame that I can then convert into a PyTorch tensor together with their corresponding target labels. Sample data from World Bank and IMF is shown in Appendix 1.

First, World Bank and IMF data extracts are loaded in via Pandas. However, as I need quarterly data and World Bank data doesn't come in quarterly, I used Cubic Spline to disaggregate the annual data into quarterly data. There are a variety of literature that shows Cubic Spline being used to disaggregate economic data such as by Rashid (2013) and Reber (2014) to name a few. I implemented this through Pandas. The inflation data is derived from the CPI data using the formula:

$$Inflation = \frac{CPI_{Current} - CPI_{Previous}}{CPI_{Previous}}$$

The columns are also arranged based on the categories as this will be useful when the CNN model is run.

The data is then cleaned by replacing the missing values by NaN and replacing the IMF countries by the Alpha-3 codes. The combined data should be in the format of Figure 1 below:

Figure 1: Combined IMF and WB Data

	Country Code	Series Name	1996 Q4	1997 Q1	1997 Q2	1997 Q3	1997 Q4	1998 Q1	1998 Q2	1998 Q3	...	2019 Q3	2019 Q4	2020 Q1
0	AGO	Agriculture, forestry, and fishing, value adde...	7.026869	6.170805	6.421647	7.478814	9.002018	10.613408	12.034404	12.920511	...	7.812613	7.882625	8.189303
1	AGO	Central government debt, total (% of GDP)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN
2	AGO	Control of Corruption: Estimate	-1.167702	-1.169295	-1.170889	-1.172483	-1.174076	-1.17567	-1.177263	-1.178857	...	-1.103711	-1.078114	-1.046928
3	AGO	Current account balance (% of GDP)	43.395691	24.019641	8.749454	-2.943636	-11.551993	-17.644539	-22.12061	-25.596385	...	9.022828	7.412294	5.571246
4	AGO	Expense (% of GDP)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	16.017531	16.684952	17.699586
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2443	ZMB	Services, value added (% of GDP)	41.543524	40.924948	40.75313	40.961617	41.471849	42.183392	43.036698	43.954715	...	54.701244	54.603759	54.476751
2444	ZMB	TBills - Yield	58.1	46.166667	30.466667	22.066667	19.233333	17.733333	20.055333	29.289	...	21.896667	22.533333	23.203333

Then, 5 sets of data will be made corresponding to the lag of 1, 4, 5, 6, and 24. For each Country and Period, there will be multiple rows of data corresponding to the lag. For each Country and Period combination, if there is at least one NaN, then it is removed from the data point. In future, other transformation such as Box-Cox or Johnson transformation may be explored as some features seems to have a different distribution even being expressed as percentage of GDP. The data is then standardized by subtracting the mean and dividing by standard deviation on a "Series Name" level.

S&P Ratings are also transformed in order to arrive at a Country, Period table. The remaining WB and IMF data will then be merged with the S&P Ratings table in order to encode the corresponding S&P rating per Country, Period combination. The final data should look like the one in Figure 2:

Figure 2: Final Data format

Country Code	Period	Rating	Lag	Agriculture, forestry, and fishing, value added (% of GDP)	Manufacturing, value added (% of GDP)	Services, value added (% of GDP)	Trade (% of GDP)	Trade in services (% of GDP)	Central government debt, total (% of GDP)
ALB	2010 Q2	B+	T-6	1.259268834	-1.326328573	-1.64866383	-0.048655007	0.833490675	0.081373078
ALB	2010 Q2	B+	T-5	1.247207215	-1.326282523	-1.629579183	-0.063642959	0.859302792	0.081181579
ALB	2010 Q2	B+	T-4	1.238360157	-1.31811167	-1.603945229	-0.071921006	0.879777557	0.080918237
ALB	2010 Q2	B+	T-3	1.238502971	-1.305696108	-1.58362708	-0.075534837	0.889311027	0.080577679
ALB	2010 Q2	B+	T-2	1.253519513	-1.293305571	-1.581167135	-0.076670312	0.882130172	0.080159292
ALB	2010 Q2	B+	T-1	1.285696786	-1.284123502	-1.603847962	-0.076964467	0.856521664	0.079671701
ALB	2010 Q3	B+	T-6	1.247207215	-1.326282523	-1.629579183	-0.063642959	0.859302792	0.081181579
ALB	2010 Q3	B+	T-5	1.238360157	-1.31811167	-1.603945229	-0.071921006	0.879777557	0.080918237
ALB	2010 Q3	B+	T-4	1.238502971	-1.305696108	-1.58362708	-0.075534837	0.889311027	0.080577679
ALB	2010 Q3	B+	T-3	1.253519513	-1.293305571	-1.581167135	-0.076670312	0.882130172	0.080159292
ALB	2010 Q3	B+	T-2	1.285696786	-1.284123502	-1.603847962	-0.076964467	0.856521664	0.079671701
ALB	2010 Q3	B+	T-1	1.328890747	-1.276036243	-1.643433308	-0.075396729	0.821377466	0.079096961
ALB	2010 Q4	B+	T-6	1.238360157	-1.31811167	-1.603945229	-0.071921006	0.879777557	0.080918237
ALB	2010 Q4	B+	T-5	1.238502971	-1.305696108	-1.58362708	-0.075534837	0.889311027	0.080577679
ALB	2010 Q4	B+	T-4	1.253519513	-1.293305571	-1.581167135	-0.076670312	0.882130172	0.080159292
ALB	2010 Q4	B+	T-3	1.285696786	-1.284123502	-1.603847962	-0.076964467	0.856521664	0.079671701
ALB	2010 Q4	B+	T-2	1.328890747	-1.276036243	-1.643433308	-0.075396729	0.821377466	0.079096961
ALB	2010 Q4	B+	T-1	1.373759638	-1.26606481	-1.687432138	-0.070152198	0.789128595	0.078429301
ALB	2011 Q1	B+	T-6	1.238502971	-1.305696108	-1.58362708	-0.075534837	0.889311027	0.080577679
ALB	2011 Q1	B+	T-5	1.253519513	-1.293305571	-1.581167135	-0.076670312	0.882130172	0.080159292
ALB	2011 Q1	B+	T-4	1.285696786	-1.284123502	-1.603847962	-0.076964467	0.856521664	0.079671701

4. Data Distribution and Feature Correlation

Initial iterations when data such as Total Reserves and Short-Term Debt were included resulted to a bias data distribution of mostly B and CCC rated countries. Hence, I made the decision to remove them, and the resulting data distribution is as the Histograms in Figure 3 and 4. Rating Mapping is also shown in Table 2 below.

As can be seen, we have abundance of data for Ratings 1 to 5 (No Modifiers) while very little data for Ratings 6 to 9. There is even no data at all for Rating 8. Similar trend can be seen on the data with Rating Modifiers. To adjust for this, weights are given on the Cross-Entropy Loss function that is inversely proportional to its frequency in the training data.

Table 2: Mapping of S&P Ratings

From	w/ Mod	w/o Mod
AAA	0	0
AA+	1	1
AA	2	1
AA-	3	1
A+	4	2
A	5	2
A-	6	2
BBB+	7	3
BBB	8	3
BBB-	9	3
BB+	10	4
BB	11	4
BB-	12	4
B+	13	5
B	14	5
B-	15	5
CCC+	16	6
CCC	17	6
CCC-	18	6
CC	19	7
C	20	8
SD	21	9

Figure 3: Data Distribution with No Modifiers

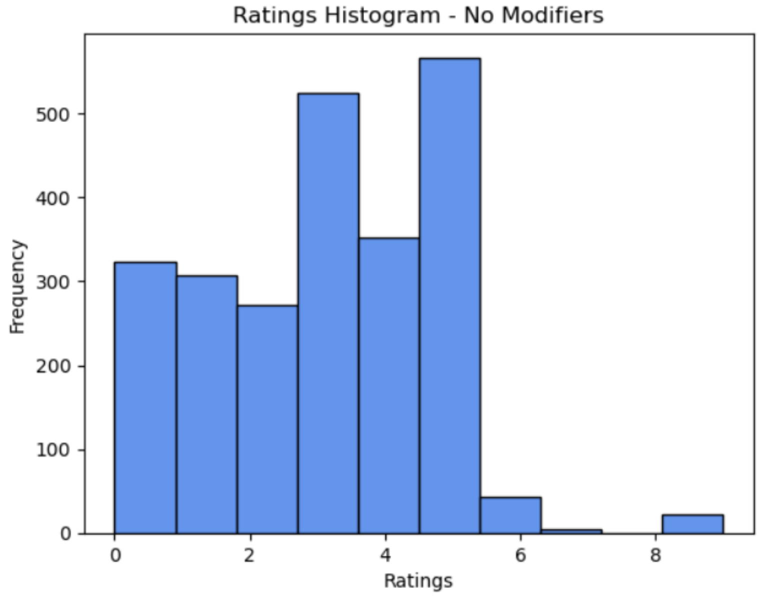


Figure 4: Data Distribution with Modifiers

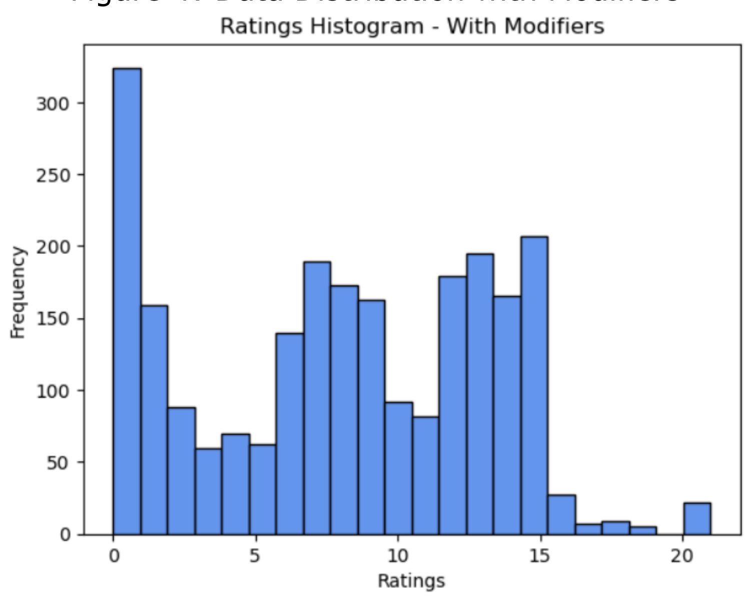


Figure 5 below shows the correlation of features. For the most part the features are not highly correlated as shown from the mostly white blocks. The highly correlated ones is the block at the lower right which is the World Bank Governance Indicators.

Figure 5: Feature Correlation Heatmap

5. Methodology

For the purpose of this project, I will be assessing the performance of the Random Forest Classifier, RNN, LSTM, CNN, and MLP based on accuracy of its prediction on the testing data. Cross-Entropy Loss will be used as the loss function for the purposes of training the data. In later models, Cross-Entropy Loss with more weights given to under represented data will be trained as well. Before ultimately choosing the below hyperparameters, several values were run initially such as hidden layers, hidden sizes, dropout probability, and others.

I will be mainly using Pandas and Numpy for data wrangling and processing. Particularly on loading, cleaning, and converting data into proper dimensions. For the models, I will be using PyTorch for the Deep Neural Networks, Scikit-Learn for Random Forest Classifier and Confusion Matrix, and Captum for Feature Attribution.

There will be two sets of training and testing data with the only difference being the target labels. One has labels without Rating Modifiers and the other one having Rating Modifiers. RNN, LSTM, and CNN will be using sequence of lagged data particularly lag of 4, 5, 6, and 24 of quarterly data. For Random Forest and MLP, I will just be using the last quarter's data. The data will be split into 75% Training Data and 25% Testing Data. The model will output values corresponding to the number of classes (10 for No Ratings Modifier, 22 for With Ratings Modifier). The class with the highest value will then be used as the predicted label. Hence, ultimately output the sovereign's Long-Term Foreign Currency Rating forecast for the next quarter.

For Sequence Models RNN and LSTM, the input will be a tensor of shape Batch Size x Sequence Size x Input Size. Table 3 below shows the hyperparameters used for this project. Visualization of the RNN and LSTM architecture is shown in Appendix 2.

Table 3: Hyperparameters for RNN and LSTM Model

RNN and LSTM	
Parameter	**Value**
Epochs	201
Learning Rate	0.001
Input Size	24
Sequence Length	[4, 5, 6, 24]
Number of Layers	2
Hidden Size	256
Number of Classes	[10, 22]
Dropout	0.075
Criterion	Cross Entropy
Optimizer	Adam
Bi-Directional	[False, True]

For CNN, the input will be a tensor of shape Batch Size x Input Channel x Sequence Size x Input Size. Table 4 shows the hyperparameters for the CNN Model. Visualization of CNN architecture is shown in Appendix 2.

Table 4: Hyperparameters for CNN Model

CNN	
Parameter	**Value**
Epochs	201
Learning Rate	0.001
Input Channel	1
Input Size	24
Sequence Length	[4, 5, 6, 24]
	In Channel = 1
	Kernel = (3,3)
	Stride = (1,1)
	Padding = (1,1)
Convolution Layer 1	Out Channel = 8
	In Channel = 8
	Kernel = (3,3)
	Stride = (1,1)
	Padding = (1,1)
Convolution Layer 2	Out Channel = 16
	Kernel = (2,2)
Pooling Layer	Stride = (2,2)
Number of Dense Layer	[1, 2]
Number of Classes	[10, 22]
Dropout	0.075
Criterion	Cross Entropy
Optimizer	SGD

For MLP, the input will be a vector of single period features. There will be two Hidden Layers with the same hidden size. Table 5 shows the hyperparameters for the MLP Model. Visualization of MLP architecture is shown in Appendix 2.

Table 5: Hyperparameters for MLP Model

MLP	
Parameter	**Value**
Epochs	201
Learning Rate	0.001
Input Size	24
Number of Layers	2
Hidden Size	[128, 256]
Number of Classes	[10, 22]
Dropout	0.075
Criterion	Cross Entropy
Optimizer	Adam
Bi-Directional	[False, True]

For Random Forest, GridSearchCV of Scikit-Learn was used to find the best number of Trees and Max Features. Table 6 below shows the hyperparameters for the Random Forest model.

Table 6: Hyperparameters for Random Forest Classifier

Random Forest		
Parameter	**w/ Modifiers**	**w/o Modifiers**
Trees	100	400
Max Features	5	5

6. Evaluation and Final Results

a. Model Results

Table 7, 8, and 9 shows the results of the trained models on different lags of data.

Table 7: RNN, LSTM, CNN results on No Rating Modifiers

No Rating Modifiers				
Model	**Lag = 4**	**Lag = 5**	**Lag = 6**	**Lag = 24**
RNN	94.42%	94.47%	94.04%	96.45%
B-RNN	93.46%	94.31%	94.21%	97.16%
LSTM	94.74%	93.66%	95.36%	96.45%
B-LSTM	95.06%	94.63%	96.03%	96.21%
CNN-1	70.65%	67.80%	74.01%	91.23%
CNN-2	69.86%	66.34%	71.03%	92.65%

Table 8: RNN, LSTM, CNN results on with Rating Modifiers

With Rating Modifiers				
Model	**Lag = 4**	**Lag = 5**	**Lag = 6**	**Lag = 24**
RNN	85.01%	86.02%	86.26%	91.71%
B-RNN	83.89%	87.32%	84.27%	90.76%
LSTM	86.12%	86.99%	86.42%	91.47%
B-LSTM	87.88%	87.64%	88.25%	92.42%
CNN-1	50.24%	47.15%	52.98%	78.91%
CNN-2	52.31%	46.18%	53.15%	81.28%

Table 9: MLP and Random Forest results

Non Temporal Methods		
Model	**No Rating Modifiers**	**With Rating Modifiers**
MLP 128 Layers	95.23%	81.07%
MLP 256 Layers	95.53%	83.16%
Random Forest	95.83%	91.37%

From the above results, Bi-Directional RNN with Lag = 24 performed the best at 97.16% on No Ratings Modifier while a Bi-Directional LSTM with Lag = 24 performed the best at 92.42% on With Ratings Modifier. Nevertheless, results of the sequential models and Non-Temporal Methods are quite close with each other. Unless further runs are done and we compute the mean and standard error of the results, it is difficult to pick the best performing classifier. Despite testing on the test set, some models may have been lucky on the split, hence showing a better result.

Non-Temporal Methods have good performance with both methods exceeding a 95% accuracy rate. Though MLP has a significantly lower accuracy rate on With Rating Modifiers, Random Forest still performed excellent and being comparable to the best results of Sequential Models.

On the other hand, CNN performed the worst. It is only until the lag was increased to 24 when its results are comparable to other models.

b. Confusion Matrices

However, accuracy rate doesn't tell the whole story of the results. Next, looking at the confusion matrix of the sequential models at Lag = 6 and the Random Forest classifier:

Figure 6: Confusion Matrix of RNN at Lag = 6

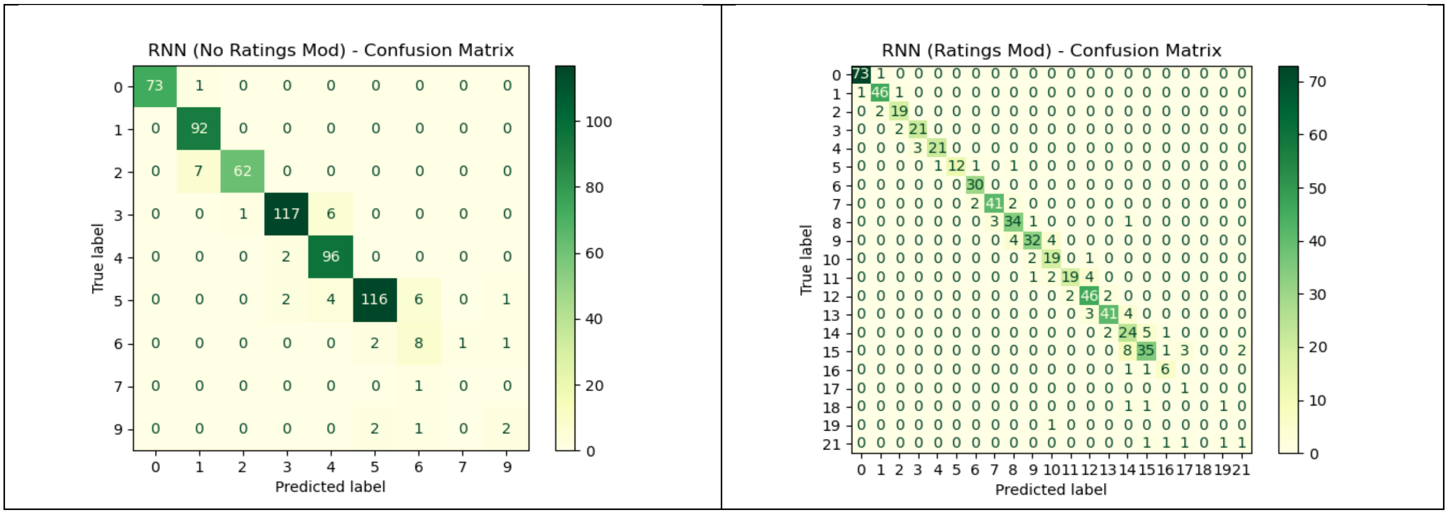


Figure 10 above shows the confusion on RNN on both No Ratings Modifier and with Ratings Modifier. The results are mostly good with correct predictions and incorrect predictions being close to the true label (+1 and -1 from the true label). This is mostly true on ratings 0 to 5 for No Ratings Modifier and ratings 0 to 15 for Ratings Modifier. We see that starting on the lower ratings, predictions are scattered, and accuracy rate severely dropped. Figure 3 and 4 supports this result as we have very limited data on these ratings. We will try to at least address this later using weights on the Cross-Entropy Loss Function.

Similarly, Figure 7 shows the Confusion Matrix for LSTM. We can see similar results with RNN.

Figure 7: Confusion Matrix of LSTM at Lag = 6

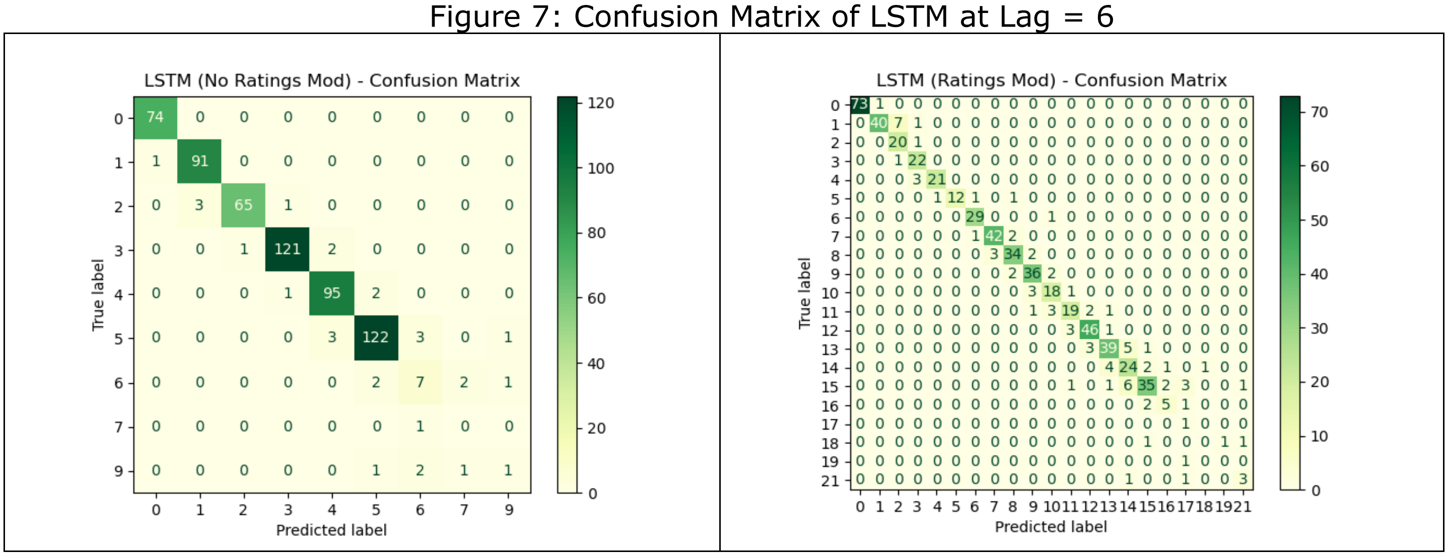
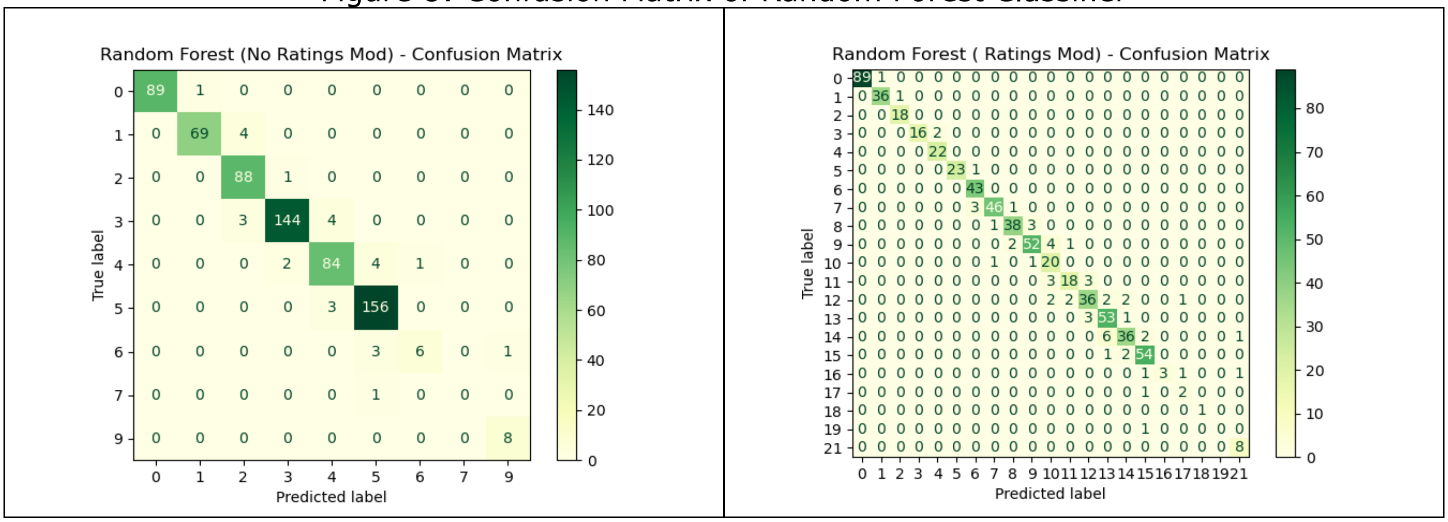


Figure 8 now shows the Confusion Matrix for the Random Forest classifier. Compared to the sequential models, the Random Forest model performed better on the lower rating grades. Again, unless further experimentation is done, it is inconclusive. This could be attributed to chance.

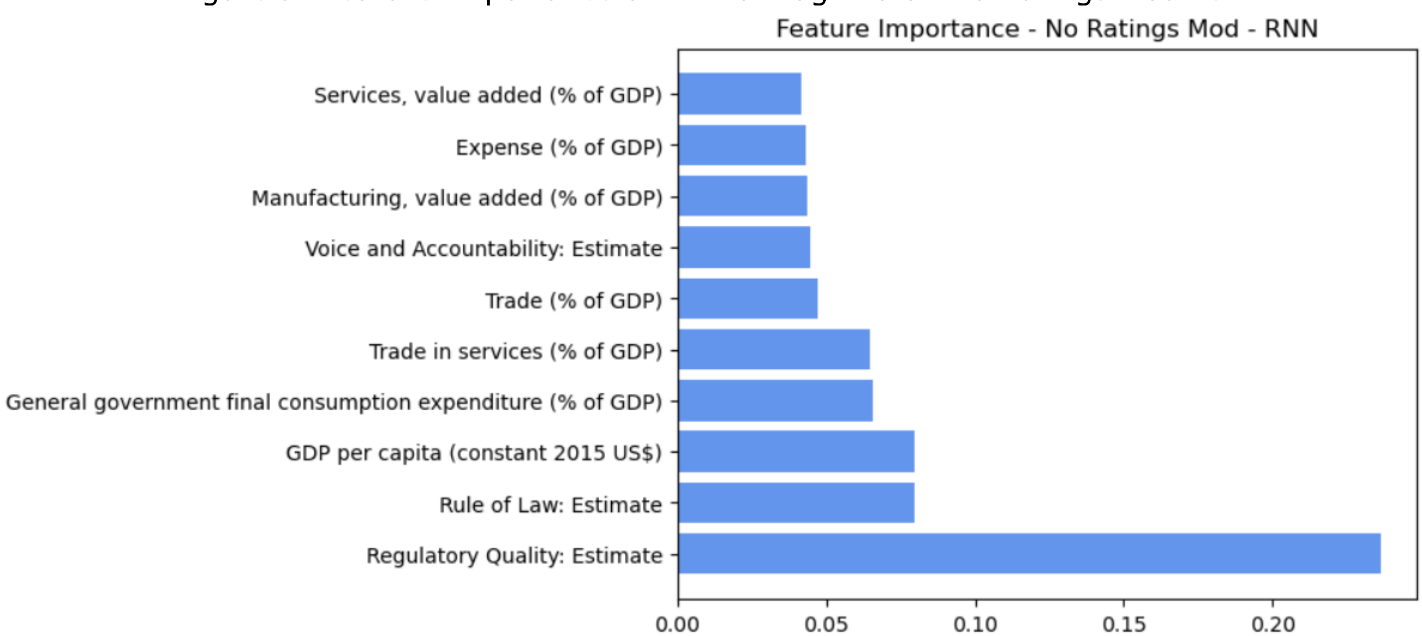
Figure 8: Confusion Matrix of Random Forest Classifier



c. Feature Importance

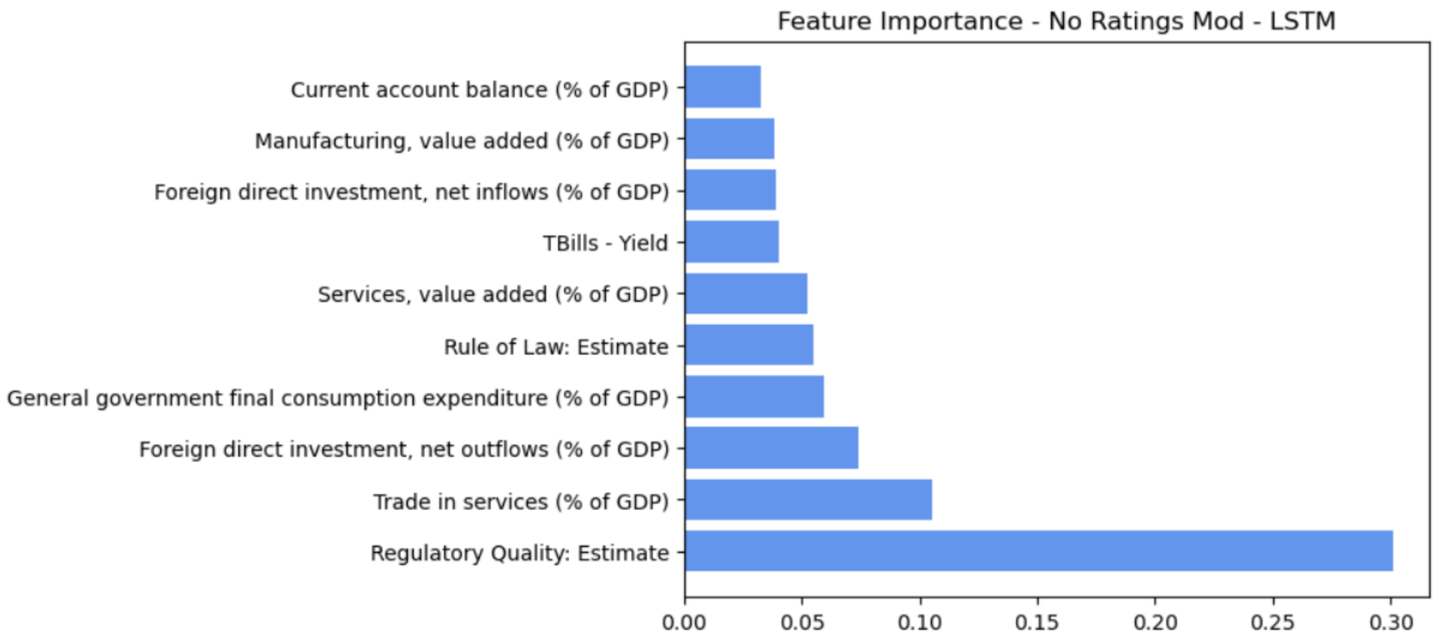
Now, let us look at the feature importance of Sequential Models and Random Forest. RNN and LSTM feature importance is calculated using Integrated Gradients of the Captum package while for Random Forest, I just used the native Feature Importance attribute of the model in Scikit-Learn. Figure 9 shows the feature importance of RNN at Lag = 6 on no ratings modifier.

Figure 9: Feature Importance of RNN at Lag = 6 on No Ratings Modifier



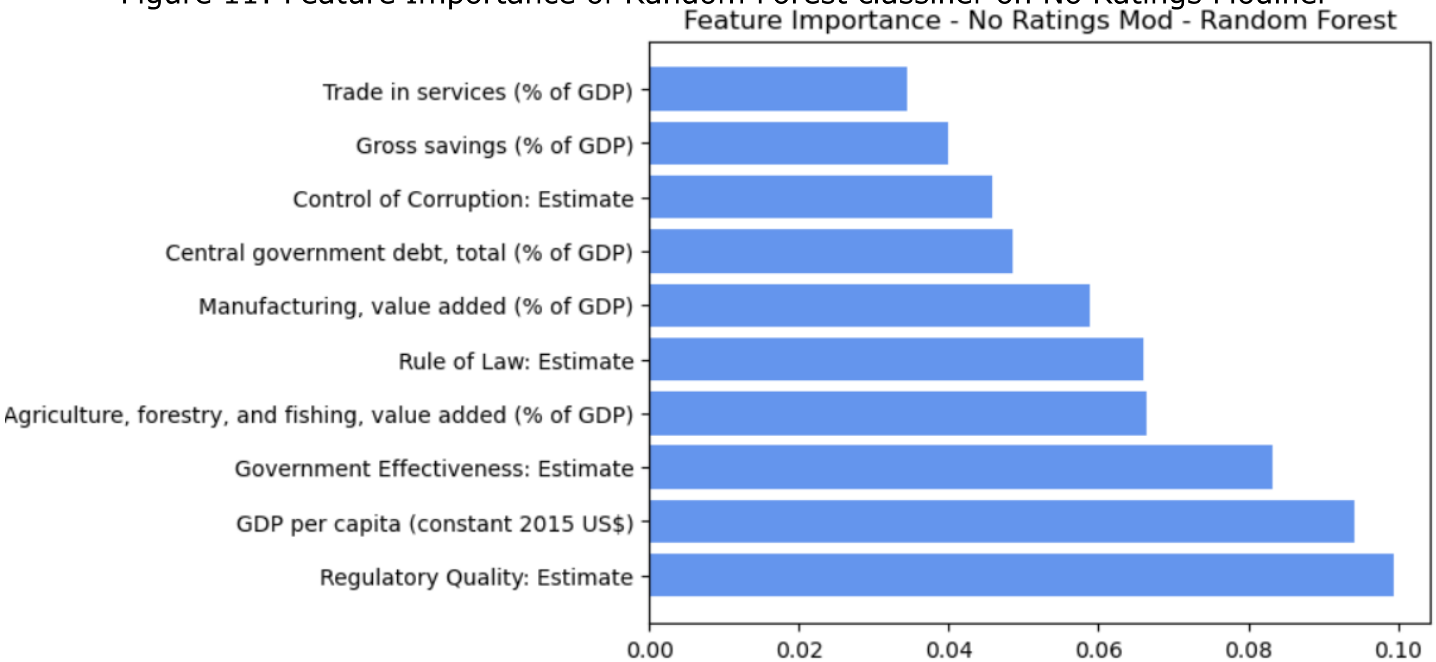
From the result, Regulatory Quality: Estimate which is a Governance Indicator, seems to be one of the most important predictor of ratings. Followed by GDP related features and other governance indicators and lastly measure of economy concentration.

Figure 10: Feature Importance of LSTM at Lag = 6 on No Ratings Modifier



From Figure 10, LSTM follows the similar pattern on RNN feature importance with Regulatory Quality leading by a wide margin followed by GDP related features, governance indicators, and economy concentration.

Figure 11: Feature Importance of Random Forest classifier on No Ratings Modifier



Like the Sequential Models, Random Forest also considers Regulatory Quality as the top predictor but not by a wide margin. It is then followed by GDP per capita.

d. Weighted Cross-Entropy Loss

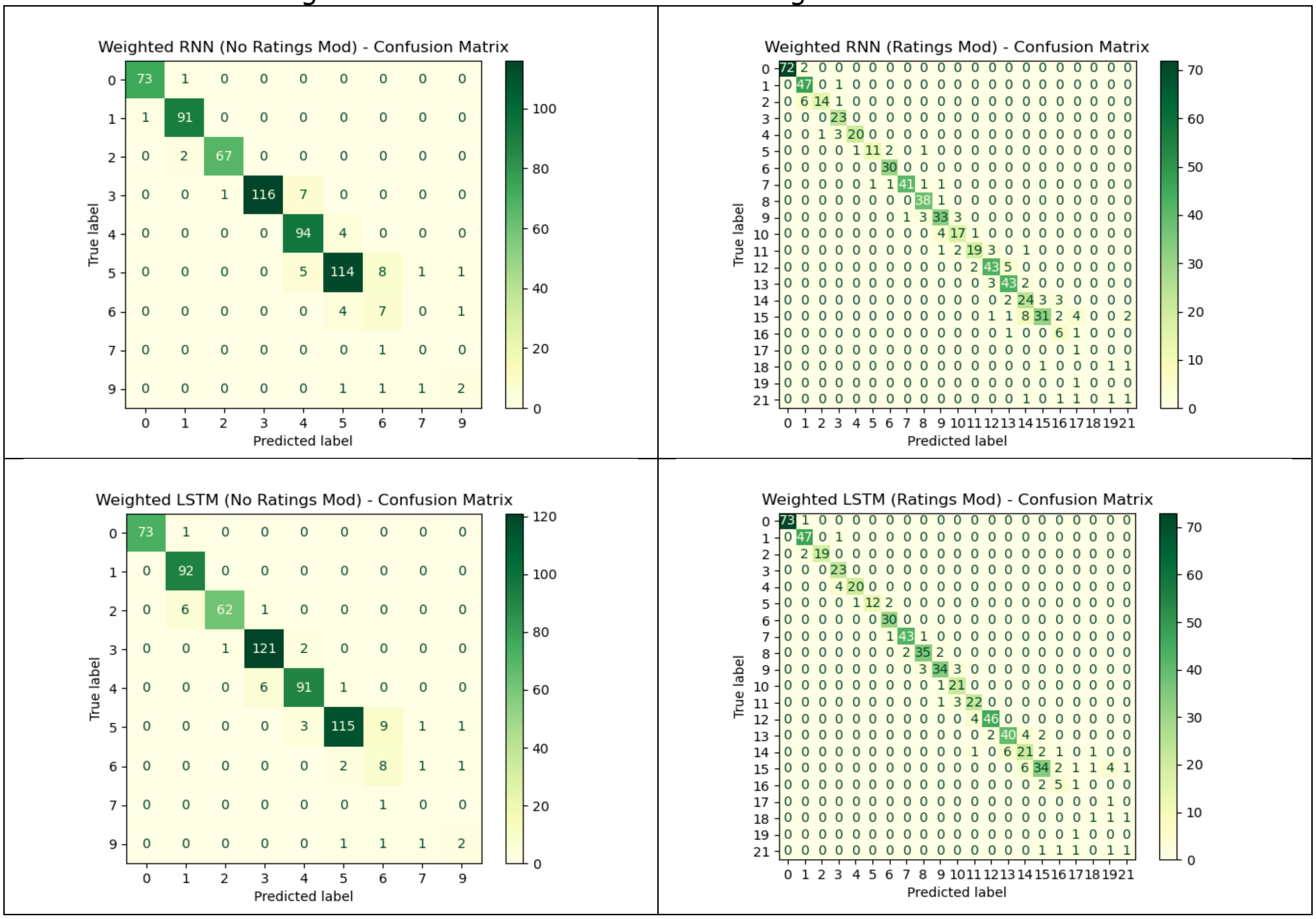
One problem encountered in earlier experiments was the limited data on the lower ratings. To address this, I used a weighted Cross-Entropy Loss in trading the Sequential Models on Lag = 6 and No Rating Modifiers. The weight w_c for each class c is given by the equation:

$$w_c = \frac{\sum_{i=0}^{C} n_i}{n_c}$$

Where C is the total number of classes and n_c is the number of instances of class c on the training data.

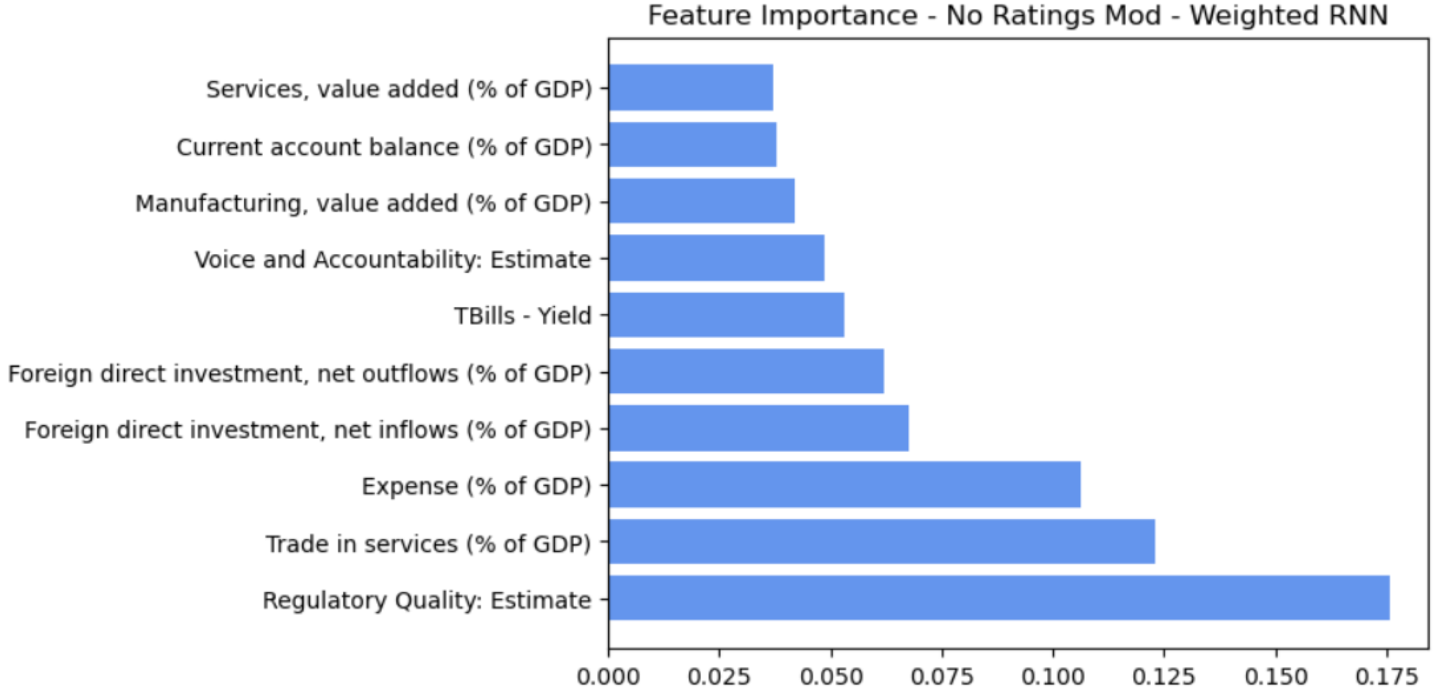
The resulting Confusion Matrices are shown in Figure 12. Despite the weighted cross-entropy, the confusion matrix on the lower rating has only improved by a little bit.

Figure 12: Confusion Matrices of Weighted RNN and LSTM



Shown in Figure 13 is the feature importance of the Weighted RNN. This shows that the margin between the feature importance has narrowed down but Regulatory Quality still leads.

Figure 13: Feature Importance for Weighted RNN on Lag = 6 and No Ratings Modifier



e. Deep Dive on the Results

If we further look at the incorrect predictions of LSTM and its ratings around that period, we see that there is usually a lag in the ratings whether that is on the side of the prediction or the actual ratings. For example, Azerbaijan on 2011 Q4 was predicted to stay at 4 while for S&P it has already upgraded to 3. On the other hand, Bahrain was predicted to be downgraded to 3 at 2010 Q4 but it is only the next quarter where it was downgraded by S&P.

Table 10: Sample comparison of Actual and Predicted Ratings for LSTM

Country Code	Period	Rating	Rating - Pred	Check
AZE	2010 Q2	4	3	FALSE
AZE	2010 Q3	4	4	TRUE
AZE	2011 Q3	4	4	TRUE
AZE	2011 Q4	3	4	FALSE
AZE	2012 Q1	3	3	TRUE
BHR	2010 Q3	2	2	TRUE
BHR	2010 Q4	2	3	FALSE
BHR	2011 Q1	3	3	TRUE
BLZ	2006 Q1	6	6	TRUE
BLZ	2006 Q2	6	7	FALSE
BLZ	2006 Q3	7	7	TRUE
BLZ	2006 Q4	9	7	FALSE
BLZ	2007 Q1	5	5	TRUE
BLZ	2011 Q4	5	5	TRUE
BLZ	2012 Q1	6	5	FALSE
BLZ	2012 Q2	6	6	TRUE

Table 11 shows the same data but for the Random Forest classifier. Similar observations can be made. Bahrain was predicted to be downgraded to 4 but it was reflected by S&P the next quarter. Same for 2002 Q3 Belize and 2014 Q2 Barbados where the ratings converged after 2 quarters. The blip in this example is Spain wherein it was rated 3 for a moment against S&P's 2.

Table 11: Sample comparison of Actual and Predicted Ratings for Random Forest

Country Code	Period	Rating	Rating - Pred	Check
BHR	2015 Q1	3	3	TRUE
BHR	2015 Q4	3	4	FALSE
BHR	2016 Q1	4	4	TRUE
BLZ	2002 Q2	4	4	TRUE
BLZ	2002 Q3	4	5	FALSE
BLZ	2002 Q4	5	5	TRUE
BLZ	2016 Q3	5	5	TRUE
BLZ	2016 Q4	7	5	FALSE
BLZ	2017 Q1	5	5	TRUE
BRB	2014 Q1	4	4	TRUE
BRB	2014 Q2	4	5	FALSE
BRB	2014 Q3	4	5	FALSE
BRB	2014 Q4	5	5	TRUE
CAN	2002 Q2	1	1	TRUE
CAN	2002 Q3	0	1	FALSE
CAN	2002 Q4	0	0	TRUE
ESP	2020 Q2	2	2	TRUE
ESP	2020 Q3	2	3	FALSE
ESP	2020 Q4	2	2	TRUE
GRC	2009 Q3	2	2	TRUE
GRC	2009 Q4	3	2	FALSE
GRC	2010 Q1	3	3	TRUE
GRC	2010 Q2	4	3	FALSE
GRC	2010 Q3	4	4	TRUE

Further study needs to be made if there are certain patterns that can be extracted from these results.

7. Conclusion

In this project we have explored the usage of Sequential Models, CNN, MLP, and Random Forest Classifier on forecasting the next quarter Foreign Currency Credit Rating. We saw that the non-temporal methods such as the MLP and Random Forest have comparable performances to the temporal methods like RNN and LSTM. The confusion matrices showed that even incorrect predictions are distributed mostly 1 rating from the actual rating. Across all models, Regulatory Control was deemed to be the most important predictor. We also dove into some actual examples of incorrect predictions and saw that the incorrect prediction can go both ways in the sense that S&P rating catches up to the predicted rating after a quarter or the predicted rating is delayed on the rating change.

8. Recommendations for Future Work

Here I'll summarize points for improvement I mentioned across the report.

1. Other relevant variables can also be explored within and outside the IMF and World Bank database. Particularly, categorical variables may be a useful addition to the model.

2. Data imputation techniques can be applied to some variables, missing or not if it is thought of to be a useful predictor. Or much better, if they can be sourced somewhere.

3. Better data transformation on some non-normally distributed variables. Some ideas might be Box-Cox or Johnson transformation.

4. Gather more data for underrepresented classes especially Rating 8 wherein there are no samples.

5. Hyperparameter tuning packages may be used such as Ray Tune or Optuna to find the optimal hyperparameters for the Deep Neural Networks.

6. Deep dive into the differences in actual and predicted ratings and see if there are patterns. Perhaps perform unsupervised learning techniques such as GMM or Kernel Mixture Models.

7. Whether new variables are added or not, feature selection can be done to reduce the input size of the models.

9. Links

[1] https://data.imf.org/?sk=4c514d48-b6ba-49ed-8ab9-52b0c1a0179b

[2] https://databank.worldbank.org/source/world-development-indicators

10.References

León-Soriano, R., Muñoz-Torres, M.J. (2012). Using Neural Networks to Model Sovereign Credit Ratings: Application to the European Union. In: Engemann, K.J., Gil-Lafuente, A.M., Merigó, J.M. (eds) Modeling and Simulation in Engineering, Economics and Management. MS 2012. Lecture Notes in Business Information Processing, vol 115. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30433-0_3

Overes, B.H.L., van der Wel, M. Modelling Sovereign Credit Ratings: Evaluating the Accuracy and Driving Factors using Machine Learning Techniques. *Comput Econ* **61**, 1273–1303 (2023). https://doi.org/10.1007/s10614-022-10245-7

Florescu, I., Golbayani, P., Wang D. Application of Deep Neural Networks to assess corporate Credit Rating. https://doi.org/10.48550/arXiv.2003.02334

Rashid, Abdul, Zanaib, Jehan. (2013). Derivation of Quarterly GDP, Investment Spending, and Government Expenditure Figures from Annual Data: The Case of Pakistan. https://mpra.ub.uni-muenchen.de/46937/1/MPRA_paper_46937.pdf

Reber, Ricci L., Pack, Sarah J. (2014). Methods of Temporal Disaggregation for Estimating Output of the Insurance Industry. https://www.bea.gov/index.php/system/files/papers/WP2014-11.pdf

Appendix 1

Figure A1.1 Sample World Bank Data

Country Name	Country Code	Series Name	Series Code	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Afghanistan	AFG	Agriculture, forestry, and fishing, value add	NV.AGR.TOTL.ZS	..	..	..	..	..	..	38.62789	37.41886	29.72107	31.11485	28.63597	30.10501	24.89227	29.2975	26.21007
Afghanistan	AFG	Central government debt, total (% of GDP)	GC.DOD.TOTL.GD	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Afghanistan	AFG	Current account balance (% of GDP)	BN.CAB.XOKA.GD	..	..	..	..	..	..	..	..	..	..	..	..	-2.32567	2.284092	-3.69524
Afghanistan	AFG	Expense (% of GDP)	GC.XPN.TOTL.GD.	..	..	..	..	..	..	..	..	..	..	20.57817	24.24326	50.7193	44.31784	50.863
Afghanistan	AFG	Exports of goods and services (% of GDP)	NE.EXP.GNFS.ZS	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Afghanistan	AFG	External balance on goods and services (% o	NE.RSB.GNFS.ZS	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Afghanistan	AFG	Foreign direct investment, net inflows (% of	BX.KLT.DINV.WD.	..	..	..	..	..	..	1.297274	1.273269	3.579894	4.352575	3.413957	1.942101	0.44912	0.461604	1.220266
Afghanistan	AFG	Foreign direct investment, net outflows (%	BM.KLT.DINV.WD	..	..	..	..	..	..	..	0.022029	-0.01341	0.024092	..	..	-0.01871	0.002033	-0.00798
Afghanistan	AFG	GDP per capita (constant 2015 US$)	NY.GDP.PCAP.KD	..	..	..	..	..	..	359.7661	363.1013	354.0337	379.9556	384.0781	429.4123	437.4198	512.409	569.2824
Afghanistan	AFG	General government final consumption exp	NE.CON.GOVT.ZS	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Afghanistan	AFG	Gross savings (% of GDP)	NY.GNS.ICTR.ZS	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Afghanistan	AFG	Imports of goods and services (% of GDP)	NE.IMP.GNFS.ZS	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Afghanistan	AFG	Manufacturing, value added (% of GDP)	NV.IND.MANF.ZS	..	..	..	..	..	..	18.82275	16.92387	17.55401	16.59821	16.38554	17.74731	17.83912	13.14988	12.52258
Afghanistan	AFG	Services, value added (% of GDP)	NV.SRV.TOTL.ZS	..	..	..	..	..	..	36.15115	37.4448	41.1109	39.00779	39.83102	40.29469	45.40983	45.24443	48.87938
Afghanistan	AFG	Trade (% of GDP)	NE.TRD.GNFS.ZS	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Afghanistan	AFG	Trade in services (% of GDP)	BG.GSR.NFSV.GD	..	..	..	..	..	..	..	..	..	..	..	..	19.93472	21.0042	20.55414
Albania	ALB	Agriculture, forestry, and fishing, value add	NV.AGR.TOTL.ZS	36.41086	31.54344	28.78565	25.91875	24.51541	22.71616	22.02511	21.97826	20.53749	18.84531	17.70835	17.15362	16.83705	16.79438	17.95587
Albania	ALB	Central government debt, total (% of GDP)	GC.DOD.TOTL.GD	37.48106	53.10782	55.5657	..	..	..	..	..	..	..	..	..	..	..	..
Albania	ALB	Current account balance (% of GDP)	BN.CAB.XOKA.GD	-3.3535	-12.0536	-2.55581	-4.83792	-4.49092	-5.54295	-9.37244	-7.24967	-4.98178	-7.09724	-7.54158	-10.7781	-15.6285	-15.3805	-11.3659

Figure A1.2 Sample IMF Data

External Sector selected indicators

Financial, Interest Rates, Government Securities, Treasury Bills, Percent per annum

Source: International Financial

Country	Scale	Base Year	1996Q1	1996Q2	1996Q3	1996Q4	1997Q1	1997Q2	1997Q3	1997Q4	1998Q1	1998Q2	1998Q3	1998Q4	1999Q1	1999Q2	1999Q3	1999Q4
Albania	**Units**		14.66	16.88	19.26	20.42	24.25	31.73	37.96	36.43	32.53	29.63	25.37	22.43	19.53	18.61	16.87	15.13
Algeria	**Units**		...	...	...	...	...	...	...	...	...	9.50	9.71	10.04	10.03	10.04	10.05	10.06
Angola	**Units**		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Antigua and Barbuda	**Units**		7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00
Armenia, Rep. of	**Units**		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Australia	**Units**		7.37	7.36	6.99	6.34	5.81	5.60	4.88	4.87	4.90	4.88	4.92	4.66	4.66	4.66	4.73	4.97
Azerbaijan, Rep. of	**Units**		...	...	...	...	...	12.45	12.00	...	12.00	...	15.20	15.10	15.37	18.12	21.63	18.10
Bahamas, The	**Units**		4.77	4.59	4.20	4.25	4.31	4.39	4.24	4.45	4.30	4.16	3.68	3.24	2.84	2.66	1.46	0.93
Bahrain, Kingdom of	**Units**		5.43	5.50	5.53	5.50	5.50	5.77	5.70	5.77	5.53	5.60	5.60	5.37	5.24	5.25	5.49	5.88
Bangladesh	**Units**		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Barbados	**Units**		8.24	7.25	6.12	5.79	5.24	3.85	1.66	3.67	5.35	5.80	5.61	5.68	5.67	5.80	5.92	5.94
Belgium	**Units**		3.37	3.23	3.15	3.02	3.21	3.22	3.50	3.60	3.53	3.61	3.49	3.40	3.01	2.51	2.51	2.83
Belize	**Units**		3.94	3.85	3.68	3.65	3.62	3.55	3.48	3.41	3.42	3.42	3.41	5.08	5.91	5.91	5.91	5.91
Bolivia	**Units**		23.88	20.61	19.65	15.60	14.68	14.64	13.40	11.88	11.82	12.41	12.61	12.49	13.88	15.31	14.37	12.70

Appendix 2

Figure A2.1 Uni-Directional 2-Layer RNN Architecture at Lag of 6 Quarters

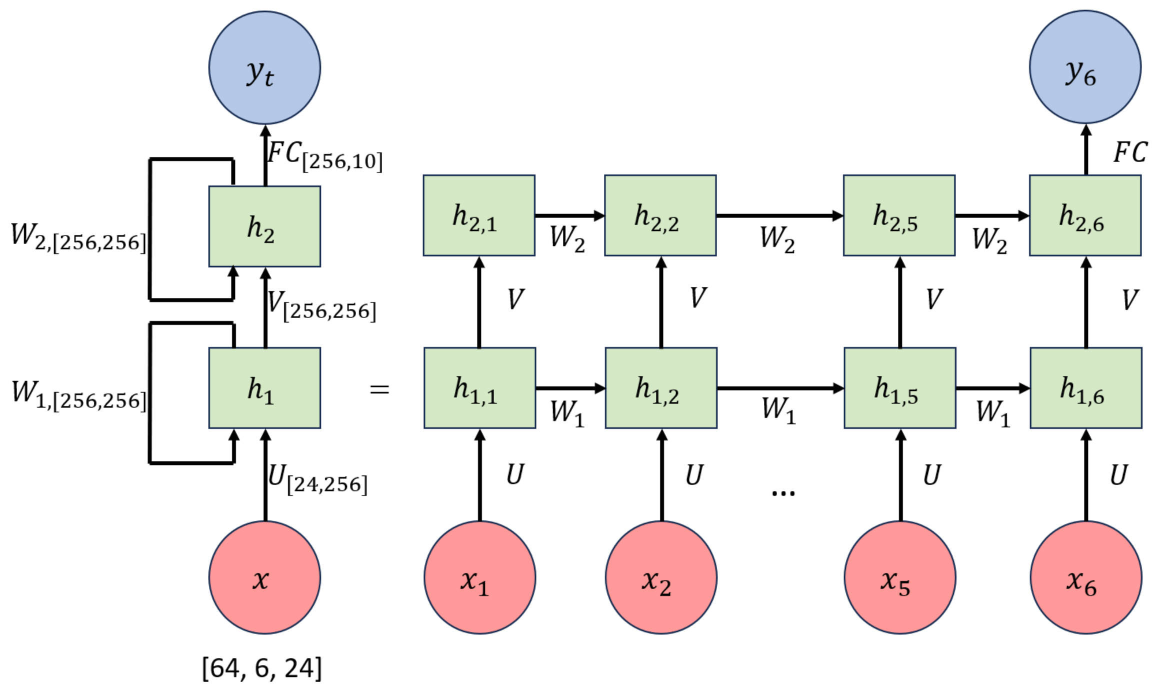


Figure A2.2 Uni-Directional 2-Layer LSTM Architecture at Lag of 6 Quarters

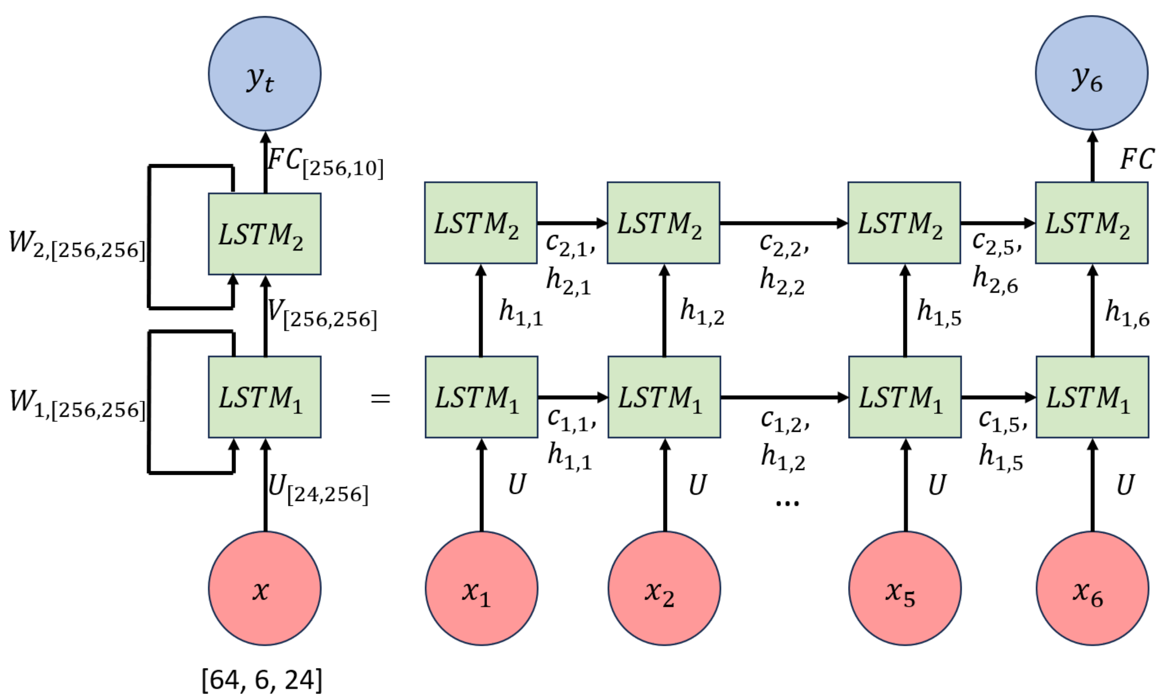


Figure A2.3 A Simple LSTM Cell

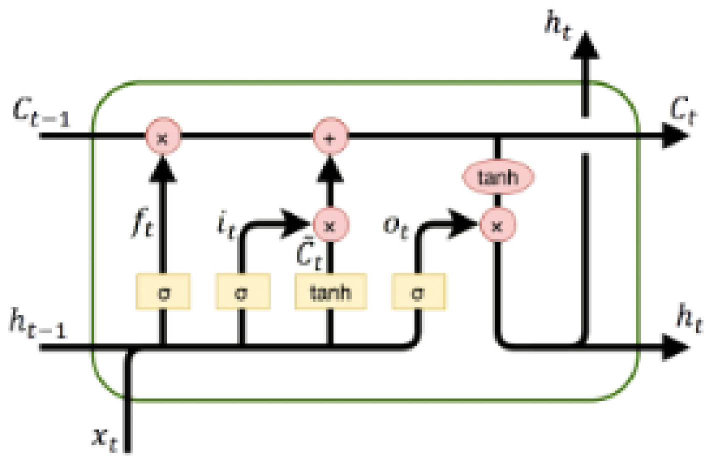


Figure A2.4 One channel, Two convolution layer, two pooling layer CNN

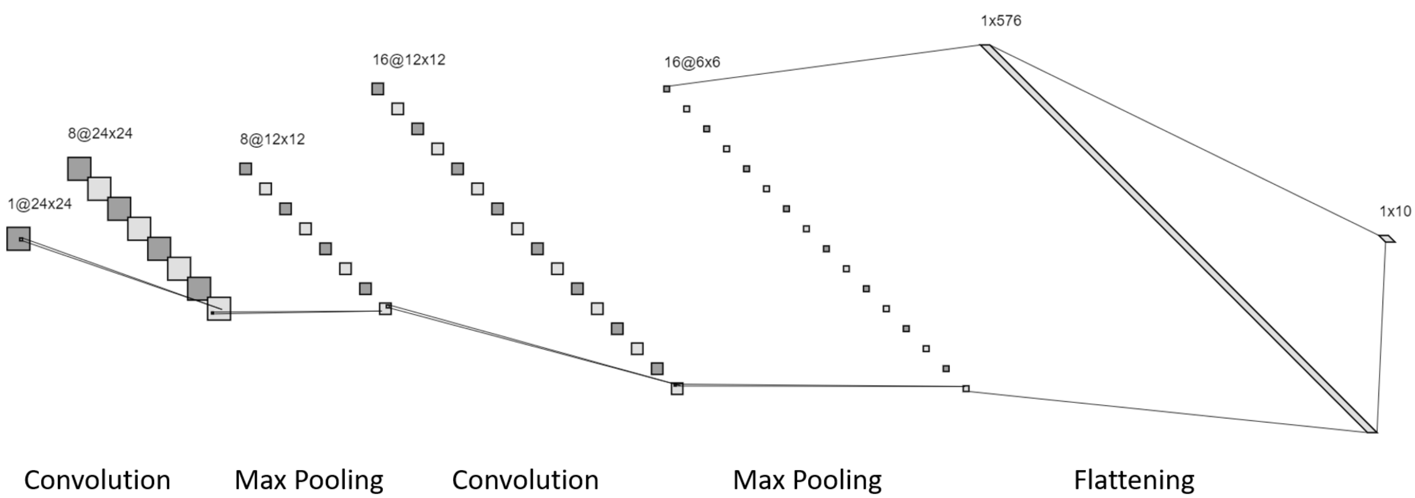


Figure A2.5 2-Layer Fully-Connected MLP with 128 hidden size